DOCUMENT RESUME

ED 388 726 TM 024 394

AUTHOR	Rock, Donald A.; And Others
TITLE	Psychometric Report for the NELS:88 Base Year through Second Follow-Up. National Education Longitudinal Study of 1988. Statistical Analysis Report.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-95-382
PUB DATE	Aug 95
NOTE	172p.; For a related document, see ED 334 241.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Statistical Data (110)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Academic Achievement; Achievement Gains; *Bayesian Statistics; Citizenship Education; *Cognitive Tests; Difficulty Level; Geography; Grade 8; Grade 12; History; *Longitudinal Studies; Mathematics Achievement; National Surveys; Reading Comprehension; Research Methodology; *Scaling; Scores; Secondary Education; *Test Construction; Test Items
IDENTIFIERS	Ceiling Effects; Item Parameters; *National Education Longitudinal Study 1988; Test Specifications

ABSTRACT

This report documents the development and validation of the National Education Longitudinal Study of 1988 (NELS:88) cognitive test battery. The cognitive test battery assesses growth between grades 8 and 12 in the content areas of reading comprehension, mathematics, science, and history/citizenship/geography. The battery was part of the NELS:88 administered to a representative sample of eighth graders in 1988 and again in 1990 and 1992. To minimize the floor and ceiling effects that typically distort gain scores, special procedures were designed into the battery. It used a two-stage multilevel procedure that attempted to tailor the difficulty of the test items to the performance level of a particular student through the administration of different test forms. Special vertical scaling procedures were used that allowed for Bayesian priors on subpopulations for both item parameters and scale scores. This report documents the test specifications for the multilevel forms and the scaling procedures, and it compares the more traditional non-Bayesian approaches to scaling longitudinal measures with the Bayesian approach. The multilevel approach was found to increase accuracy of measurement and to reduce floor and ceiling effects. (Contains 30 tables, 3 figures, 7 appendixes of supplemental information, and 42 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 388 726

NATIONAL CENTER FOR EDUCATION STATISTICS

Statistical Analysis Report August 1995

National Education Longitudinal Study of 1988

Psychometric Report for the NELS:88 Base Year Through Second Follow-Up

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U.S. Department of Education
Office of Educational Research and Improvement NCES 95-382

TM024394

BEST COPY AVAILABLE

NATIONAL CENTER FOR EDUCATION STATISTICS

Statistical Analysis Report August 1995

National Education Longitudinal Study of 1988

Psychometric Report for the NELS:88 Base Year Through Second Follow-Up

Donald A. Rock
Judith M. Pollack
Educational Testing Service

Peggy Quinn, Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement NCES 95-382

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Jeanne E. Griffith
Acting Commissioner

National Center for Education Statistics

The purpose of the Center is to collect and report "statistics and information showing the condition and progress of education in the United States and other nations in order to promote and accelerate the improvement of American education."—Section 402(b) of the National Education Statistics Act of 1994 (20 U.S.C. 9001).

August 1995

Contact:
Peggy Quinn
(202) 219–1743

Executive Summary

This report documents the development and validation of the NELS:88 cognitive test battery. The cognitive test battery assesses longitudinal growth between grades 8 and 12 in four content areas - reading comprehension, mathematics, science and history/citizenship/geography. The cognitive battery was part of the larger National Education Longitudinal Study of 1988 that was monitored by the Longitudinal and Household Studies Branch (LHSB) of the National Center for Education Statistics (NCES). The NELS:88 test battery was administered to a representative sample of 8th graders in the spring of 1988, who were then retested in the spring of 1990 and 1992. Response rates varied between 93 to 96 percent for the in-school 8th and 10th graders and dropped to about 81 percent for the twelfth graders. There was some tendency for students from low socio-economic backgrounds to be over-represented among the non-respondents.

In order to minimize floor and ceiling effects which typically distort gain scores, special procedures were designed into the development and administration of the cognitive test battery. The test battery used a two-stage multilevel procedure that attempted to tailor the difficulty of the test items to the performance level of a particular student. For example, students who performed very well on their 8th grade mathematics test received a relatively more difficult form in tenth grade than those scoring in the middle or in the lower range on their 8th grade test. There were three forms varying in difficulty in mathematics and two in the reading area in both grades 10 and 12. Since tenth and twelfth graders were taking forms that were more appropriate for their level of ability/achievement, measurement accuracy was enhanced and floor and ceiling effects could be minimized. The remaining two content areas, science and history/citizenship/geography were only designed to be grade level adaptive i.e., have a different form for each grade, and therefore did not have multiple forms varying in difficulty within grade.

In order to maximize the gain from using an adaptive procedure, special vertical scaling procedures were used that allow for Bayesian priors on subpopulations for both item parameters and scale scores. This report documents the test specifications for the multilevel forms as well as the Bayesian procedures used in the vertical scaling. The report also includes a comparison of more traditional non-Bayesian approaches to scaling longitudinal measures with the Bayesian approach.

It was found that the multilevel approach did increase the accuracy of the measurement, and when used in combination with the Bayesian item parameter estimation, reduced floor and ceiling effects when compared to the more traditional item response theory approaches.

Acknowledgments

The authors would like to thank the test development staff at the Educational Testing Service for their exemplary work in developing the item pools, especially Kalle Gerritz, Trudy Conlan, Beth Brownstein, Mary Gribben, and Eve Niedergang. In addition the authors would like to thank the NELS:88 Project Director, Steven Ingels, and his staff at NORC, as well as Jeff Owings (Branch Chief) and Peggy Quinn (Project Officer) at NCES for their help in this cooperative effort. We would especially like to thank them for their patience and support in allowing us to experiment with new approaches to vertical scaling. We also would like to thank Ron Hambleton of the University of Massachusetts and Kentaro Yamamoto of ETS for their very wise counsel on numerous technical matters. Finally, we would like to thank reviewers of earlier drafts of this report for their helpful comments. Specifically, we are indebted to other reviewers who provided helpful comments and advice on an earlier draft including Bob Burton, Ralph Lee, and Steve Gorman of NCES, Mark Reckase at ACT, Jim Crouse of the University of Delaware, and Tim Madigan at the Census Bureau.

Table of Contents

Page

Executive Summary iii

Acknowledgments iv

List of Tables vi

List of Figures viii

List of Appendices viii

Chapter 1: Introduction 1

Sample and Completion Rates 1

Chapter 2: NELS Test Specifications 3

Aims and Objectives 3
Two Stage Testing in a Longitudinal Framework 5
Specifications for Individual Tests 6
Matching Test Content to Curriculum 7
Reading 9
Mathematics 16
Science 26
History/Citizenship/Geography 30

Chapter 3: IRT Scaling for Longitudinal Measurement and Equating to Earlier Cohorts 37

Differential Item Functioning (DIF) 45
Speededness 47
Motivation 51

Chapter 4: Normative and Proficiency Level Scores 59

IRT Estimated Number Right 59
IRT Theta "T" Score 60
Cross-Sectional Scores 60
Criterion Referenced Proficiency Scores 61
General Description of the Proficiency Levels 61
NAEP Equated Score 64

Chapter 5: Psychometric Properties of the NELS:88 Scores 67

Reliability of the IRT Scores 67
Construct Validity of the NELS:88 Content Areas 68

References 85

List of Tables

		Page
Table 1.1:	Proportion of the Core Panel Sample Participants with All Tests On All Occasions	2
Table 2.1:	NELS:88 Reading Specifications Content by Process by Test Forms	10
Table 2.2:	Reading: Proportion Correct	11
Table 2.3:	Reading: R-Biserial	14
Table 2.4:	NELS:88 Math Specifications Content by Process by Test Forms	17
Table 2.5:	Math: Proportion Correct	18
Table 2.6:	Math: R-Biserial	22
Table 2.7:	NELS:88 Science Specifications Content by Process by Test Forms	26
Table 2.8:	Science: Proportion Correct	27
Table 2.9:	Science: R-Biserial	29
Table 2.10:	NELS:88 History Specifications Content by Test Forms	30
Table 2.11:	History/Citizen/Geography: Proportion Correct	31
Table 2.12:	History/Citizenship/Geography: R-Biserial	33
Table 3.1:	Means, Standard Deviations and Ranges of IRT Parameters	44
Table 3.2:	Counts of "C" Level DIF Items	48
Table 3.3:	Percentages of Selected Subgroups Who Attempted the Last Item for Each Cognitive Test	49
Table 3.4:	Percentage of Subgroups with Scorable Tests, Unweighted	51
Table 3.5:	Percentage of Subgroups with Scorable Tests, Weighted	54
Table 4.1:	Comparison of the NAEP and NELS Twelfth Grade Samples	64
Table 5.1:	Reliability of Theta	67
Table 5.2:	Intercorrelations of Content Areas Within and Across Administrations	69
Table 5.3:	Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math Correlations of Gains and Grade 12 Status with Background Variables	70
Table 5.4:	Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math Correlations of Gains and Grade 12 Status with Math Courses Taken	72
Table 5.5:	Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math Correlations of Gain with Initial (Grade 8) Status	74
Table 5.6:	Correlations of Background Variables with Second Follow-up Status and Gains: Reading	77
Table 5.7:	Correlations of Transcript Variables with Second Follow-up Status and Gains: Reading	79

List of Tables (Cont'd)

Table 5.8: Correlations of Background Variables with Second Follow-up Status and Gains: Science 80

Table 5.9: Correlations of Transcript Variables with Second Follow-up Status and Gains: Science 82

Table 5.10: Correlations of Background Variables with Second Follow-up Status and Gains: History/Citizenship/Geography 83

Table 5.11: Correlations of Transcript Variables with Second Follow-up Status and Gains: History/Citizenship/Geography 84

List of Figures

Page

Figure 3.1: Probability of Correct Answer 38
Figure 3.2: Items with Different Difficulty (B) 39
Figure 3.3: Items with Different Discrimination (A) 40

List of Appendices

Appendix A: Reading 89
Appendix B: Math 97
Appendix C: Science 107
Appendix D: History/Citizenship/Geography 113
Appendix E: Test Item Map 119
Appendix F: Invariance of Item Parameters Across Years 127
Appendix G: Test Information Function--Theta (Ability) 137

Chapter 1
Introduction

The National Education Longitudinal Study of 1988 (NELS:88) is designed to monitor the transition of a national sample of young adults as they progress from eighth grade to high school and then on to postsecondary education and/or the world of work. The NELS:88 surveys are monitored by the Longitudinal and Household Studies Branch (LHSB) of the National Center for Education Statistics (NCES). NELS:88 is the third and most recent in a series of longitudinal studies that are designed to provide timely information on trends in academic achievement. The two earlier longitudinal studies sponsored by NCES were the National Longitudinal Study of the high school class of 1972 (NLS-72) and the High School and Beyond (HS&B) study of 1980.

The primary purpose of the NELS:88 data collection is to provide policy relevant information concerning the effectiveness of schools, curriculum paths, special programs, variations in curriculum content and exposure, and/or mode of delivery in bringing about educational growth. In addition to the test scores described in this report, the NELS:88 database contains a great deal of data on factors relevant to cognitive growth, including student questionnaires with information on family background, aspirations and attitudes and experiences in and out of school; high school transcripts; and teacher, school and parent questionnaires. The sample was designed to provide sufficient numbers of students in "high risk" subpopulations to allow for separate analysis of the growth patterns for these critical subgroups. Given the ambitious educational achievement goals that are being set for the year 2000, it is critical that we gather evidence now on how variations in student characteristics interact with variations in the content and processes of educational programs in bringing about cognitive growth.

The purpose of this report is to document the rationale and technical decisions that were carried out in the design, development and scaling of the cognitive battery.

Sample and Completion Rates

While the base year (1988) participating sample was 24,599, a subsample was selected for follow-up in the subsequent years, with varying probabilities depending on how they clustered in schools. Panel test data were obtained on approximately 12,000 core sample individuals who had useable cognitive test data on all three (1988, 1990, 1992) occasions. In addition to the core panel sample individuals, there were augmented state and other special samples at the base year and succeeding follow-ups. Freshened samples were also added at the first and second follow-up to insure a representative sample of students within a grade. Additional details about the sample design and survey procedures may be found in the second follow-up user's manual (Ingels et al., 1994). Table 1.1 below presents the test completion rates for selected subpopulations for individuals in the core panel sample only.

Inspection of Table 1.1 indicates that approximately two thirds of the total target sample have all four cognitive scores on all three occasions. Much of the analysis in this psychometric report will be based on this panel sample. Cross-sectional (within-year) analyses that do not require data at all three time points will include students who were in the NELS:88 core sample but were not tested at all three points in time; other statistics that are internal to the tests themselves and do not make reference to national estimates may include the state augmentation samples that were not part of the NELS:88 core. These less stringent criteria lead to significantly greater participation rates than those shown in Table 1.1. More detailed discussions about non-response rates are presented in the section on motivation. A detailed discussion of sample selection and weighting procedures may be found in Ingels et al. (1994).

Table 1.1
Proportion of the Core Panel Sample Participants with All Four Cognitive Tests On All Three Occasions

	Eligible Core Panel Sample		Percentages With All Tests On All Occasions	
	RAW N	WTD N	% RAW N	% WTD N
Total	16489	2970835	70	65
Male	8140	1492789	69	66
Female	8349	1478047	70	65
Asian	995	105878	69	66
Hispanic	2017	307485	61	58
Black	1628	390455	63	52
White	11662	2122702	72	69
Public School[a]	12585	2253702	74	72
Catholic School[a]	850	149699	79	75
NAIS Private[a]	930	32107	73	74

[a] The classification by school type only includes those individuals who were enrolled in school. The remaining classifications, gender and race, includes all students whether they are enrolled or not.

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Chapter 2
NELS Test Specifications

This chapter will discuss the special considerations in testing a national sample of students in several subject areas over a four-year time span. The rationale for the design of multiple overlapping test forms is described, as well as the considerations in choosing the timing and content of each form.

Aims and Objectives

The test specifications of the NELS:88 longitudinal test battery are dictated by its primary purpose: accurate measurement of the status of **individuals** at a given point in time, as well as their growth over time. Like its predecessor, the 1980 High School and Beyond (HS&B) test battery, the National Education Longitudinal Study (NELS:88) test battery was developed to measure both individual status and growth in a number of achievement areas. The four achievement areas are Reading Comprehension, Mathematics, Science, and History/Citizenship/Geography(H/C/G). However, unlike the HS&B assessment, which was designed only to measure growth between the tenth and twelfth grades, the NELS:88 battery is designed to measure growth in achievement between the eighth, tenth and twelfth grades. Since the NELS:88 assessment spans four years with repeated testing of the same student cohort in the eighth, tenth and twelfth grades, it calls for a more flexible testing approach than was required in the HS&B longitudinal assessment.

The construction of the NELS:88 eighth grade battery is in some sense a delicate balancing act between several competing objectives. Many of these objectives were suggested by the NELS Technical Review Panel (TRP) and/or NCES project staff during the base year development. Some of these objectives were as follows:

- The NELS:88 test battery should cover four content areas - Reading, Mathematics, Science, and History/Citizenship/Geography.

- Item selection should be curriculum-relevant, with emphasis on concepts, skills and general principles. When measuring change or developmental growth, the overemphasis on isolated facts at the expense of conceptual and/or problem-solving skills may lead to distortions in the gain scores due to forgetting. More will be said about this later.

- The tests should be relatively unspeeded with the vast majority of students completing all tests.

- There should be little evidence of floor or ceiling effects.

- Reliabilities of the component tests should be psychometrically acceptable for the purpose of measuring **individual** status as well as growth. While much of the analysis using the NELS database will probably be at the group level, there will be many studies that use the test scores as covariates. In such cases the reliability of the covariates becomes important. Also when measuring change we need evidence that we are measuring the same things over time.

- The accuracy of measurement, i.e., the standard error of measurement, should be relatively constant across SES, sex and racial/ethnic groups. In fact, the NELS:88 battery was specifically designed to reduce the gap in reliabilities that is typically found between the majority group and the racial/ethnic minority groups.

- The individual test content areas should demonstrate some discriminant validity. That is, while the tests should be internally consistent and be characterized by a large dominant factor, when factor analyzed together, they should yield a relatively "clean" although oblique four factor solution. The four factors should be defined by the four content areas. The Base Year Psychometric Report (Rock & Pollack, 1991) presents results for the four factor solution. Because of the multilevel nature of two of the four tests in the tenth and twelfth grades, intercorrelations among the test scores rather than factor analysis results are presented in this report.

- Subscores and/or proficiency scores should be provided where psychometrically justified. The test specifications were designed to provide behaviorally-anchored proficiency (mastery) scores in the areas of Reading, Mathematics, and Science.

- The NELS:88 test battery should attempt to minimize Differential Item Functioning (DIF) across gender and racial/ethnic groups that arises from irrelevant content that favors one or more of the groups.

- The NELS:88 test battery should share sufficient common items both across and within grade level forms, and with the HS&B battery, to provide articulation of scores for vertical equating in NELS:88 as well as cross-sectional equating with the 1980 HS&B sophomore cohort in mathematics.

- There should be sufficient item overlap between the National Assessment of Educational Progress (NAEP) mathematics test and the twelfth grade NELS:88 mathematics test to cross-walk to the NAEP mathematics scale if desired.

- The reading test passages should provide relatively broad content coverage and have items that span at least three cognitive process areas. There also should be at least one passage that identifies in some way with minority concerns. Similarly, there should be at least one passage in which the main character is a female.

- The four content areas Reading, Mathematics, Science, and History/Citizenship/ Geography must be administered (including time for administration instructions) within one hour and a half.

- The tests should be sufficiently reliable to support change measurement, and be characterized by a sufficiently dominant underlying factor to support the Item Response Theory (IRT) model. This latter requirement is necessary to support the vertical equating between retestings as well as the cross-sectional linking with HS&B and NAEP, if desired. The IRT vertical equating puts the scores within a given content area on the same scale regardless of the grade in which the score was obtained. This allows the user to interpret scores the same way whether they were from the eight, tenth, or twelfth grade. Independent of the vertical scaling, the testing time constraints made achieving desired reliabilities problematic without introducing some sort of adaptive testing. In order to achieve this level of reliability, as well

as reduce the possibility of "floor and ceiling" effects, the Mathematics and Reading tests were designed to be multilevel at the tenth grade and twelfth grade. The multilevel adaptive approach is discussed below.

- While the NELS:88 battery provides test scores with the usual normative interpretation, it was also designed to have "mastery" level scores in mathematics, reading, and science. These multiple criterion-referenced levels serve two functions. First, they help with respect to the interpretation of what a score level "means" in terms of what Mary or Johnny can or cannot do. Second, they are useful in measuring change at particular score points along the score scale. In particular, when certain school processes can be expected to be reflected in score changes taking place at specific points along the score scale, then changes in percent or probability of mastery at that point in the scale would be better measures of the impact of the school process on student growth than would changes in the overall test score. More details about these criterion-referenced scores and their interpretation will be presented in the section on cognitive scores.

Two Stage Multilevel Testing in a Longitudinal Framework

The potentially large variation in student growth trajectories over a four year period argues for a longitudinal "tailored testing" approach to assessment. That is, in order to accurately assess a student's status both at a given point in time as well as over time, the individual tests must be capable of measuring across a broad range of ability/achievement. If the **same** test, in say, Mathematics and Reading Comprehension were administered to the same student at the eighth, tenth, and twelfth grades, the potential for observing "floor effects" at grade eight and "ceiling effects" at grade twelve is greatly increased. Of course if all four tests were quite long and included many very difficult as well as many very easy items, then theoretically there would be little opportunity for floor and ceiling effects to operate.

Unfortunately operational versions of the test must be relatively short in order to minimize the testing time burden on the students and their school systems. The solution to this problem was to use a two-stage testing procedure that allows one to at least partially tailor a test form to a particular individual's ability/achievement level.

That is, a two-stage multilevel longitudinal testing procedure was implemented that used the eighth grade reading and mathematics test results for each student to assign him or her to a different form of the test when he or she was re-tested in tenth grade. The same procedure was repeated in the twelfth grade. For example, students scoring relatively high on the eighth grade test, (top twenty-five percent) in say, mathematics were given a more difficult mathematics test form when they were retested as tenth graders. Students scoring relatively low in the eighth grade (bottom twenty-five percent) received an easier form when retested as tenth graders. Students scoring in the middle range received an "average" difficulty mathematics form. Since tenth and twelfth grade students would be taking forms that were in a sense appropriate to their particular level of ability/achievement, measurement accuracy would be enhanced, and floor and ceiling effects would be minimized. The relative absence of ceiling effects should make the assessment of gain more accurate for students who had relatively high scores as eighth graders and/or as tenth graders. Similarly, an accurate estimate of gain for low scoring eighth graders should also be enhanced, since floor effects should be minimized.

In summary, the tenth and twelfth grade mathematics and reading tests incorporated multilevel forms differing in difficulty. The tenth and twelfth grade science and history/citizenship/geography tests

were grade level adaptive in the sense that everyone took the same form within a grade but each succeeding grade level form included additional more difficult items.

What does the utilization of a two-stage multilevel procedure have to say about how the components of the NELS:88 battery should be constructed? With respect to the eighth grade, two of the eighth grade tests (reading and mathematics) were to serve as "branching" or "routing" tests, and thus ideally they should have good measurement properties throughout the test score range. That is, the test scores should provide reliable information at the high, the middle, and the low end of the test score distribution since students in these score ranges could then be routed to tests of quite different average difficulties in the tenth grade.

Because of their branching role the eighth grade reading and mathematics tests were designed with somewhat more broad band measurement properties in mind. Operationally, the goal of maintaining good measurement accuracy throughout the test score range is accomplished by building tests with a relatively rectangular frequency distribution of item difficulties, that is, equal numbers of test items at each difficulty. The typical test, however, tends to follow a normal distribution of difficulties with the majority of the items in the middle difficulty range. However, if one wished to use the base year test as not only a measure of an individual's achievement status in grade 8, but also as a routing test for assignment to tenth grade forms that vary in difficulty, then one should have a more rectangular distribution of difficulty levels.

The tenth and twelfth grade tests in reading and mathematics must include sufficient linking items both across grades as well as across forms within grade to allow both cross-sectional and vertical equating using Item Response Theory (IRT) models (Lord, 1980). In the case of the science and history/citizenship/geography (H/C/G) tests, linking items need to be present across grade forms only. In mathematics and reading the average difficulty (percent getting an item correct) of the various within-grade forms should be in the .45 to .60 range, and the distribution of the item difficulties (P+) should be more peaked than for forms that are designed to measure efficiently across a broad range of ability. The P+ values are not symmetric around .50 since in theory it is assumed that fewer students need to guess when the items are somewhat easier.

While the multilevel adaptive approach used in mathematics and reading and the grade level adaptive approach used in the science and the H/C/G tests helped in minimizing floor and ceiling effects, it was decided that more recent developments in IRT models would also be necessary to take full advantage of the adaptive nature of the NELS:88 battery. More specifically, a Bayesian procedure (Mislevy & Bock, 1989; Muraki & Bock, 1987) was used in estimating both the item parameters and the ability scores. This procedure allowed for separate prior ability distributions, thereby taking into consideration the differing ability distributions associated with the various forms used across and within grades. More details will be presented about this procedure in Chapter 3 as part of a technical discussion dealing with the special IRT estimation model that was used.

Specifications for Individual Tests

Based on simulations utilizing field test results (Rock & Pollack, 1987), ETS test development experts determined the number of test items needed to provide accurate assessment of each content area, and the time required to minimize speededness. Given that the maximum allowable testing time for eighth graders was approximately one hour and thirty minutes, including five minutes for instructions, it was decided that the time would be apportioned in the following way among the test battery components:

Reading - Twenty-one questions in twenty-one minutes.
Mathematics - Forty questions in thirty minutes.
Science - Twenty-five questions in twenty minutes.
History/Citizenship/Geography - Thirty questions in fourteen minutes.

The items that were used in the final eighth grade forms were selected from a much larger pool of items composed of items from NAEP, HS&B, the Second International Mathematics Study (SIMS), ETS test files from previous operational tests, and a pool of items specifically written for the NELS:88 Battery. The selection of items for the pre-test item pools was based on the consensus of the members of subject matter committees made up of curriculum experts.

The subject matter committees consisted of educators, teachers, and college professors specializing in middle school curricula. There was considerable personnel overlap with similar subject matter committees used in the NAEP item pool development. ETS test development specialists were in attendance and worked with their respective subject matter committees in developing the eighth, tenth and to some extent the twelfth grade assessment objectives. Once the assessment objectives were agreed upon, the subject matter committee members classified the items according to the objectives. A pool of 50 Reading items, 82 Mathematics items, 42 Science items, and 60 History/Citizenship/Geography items was selected for pretesting. Field tests were administered to eighth, tenth and twelfth graders in the Spring of 1987 (Rock & Pollack, 1987). The results of the field testing were scrutinized by additional committees of subject matter experts who suggested numerous modifications in content, format and wording of the items, as well as making judgments on content coverage. Final revisions and item selections were made by project staff on the basis of their input, and reviewed by NCES staff.

Matching Test Content to Curriculum

The question of overlap between test items and curriculum content has received increasing attention over the last ten years and evaluation methodologies have come to be dominated by the **doctrine of maximal overlap** (Frechtling, 1989). Mehrens (1984) and Cronbach (1963), however, questioned whether maximal overlap is in fact desirable except possibly in those cases where a specific program is being evaluated. Mehrens argues that a close match between curricular and test content is desirable only if one wishes to make inferences about specific objectives taught by a specific teacher to a specific school. Even if one would wish to evaluate the effects of a specific teacher in a specific class, one inference of importance is the degree to which the specific knowledge taught in that class generalizes to other relevant domains.

Nitko (1989) argues that tests designed to measure individuals and to facilitate their learning within a particular instructional context are not necessarily optimum for measuring school or program differences. Similarly Airasian & Madaus (1983) suggest that the following design variables be taken into account:

(A) The ability of tests to detect differences between groups of students.
(B) The relative representativeness of the content-behavior-process sampled by test items.
(C) The parallelism of the response formats and mental processes learned during instruction with those defined by the test tasks.
(D) The properties of the scores and the way that they will be summarized and reported.
(E) The validity of the inferences about school and program effectiveness that can be made from the test results.

Experience and practice suggests that tests are unlikely to detect differences between schools and programs when total test scores are used and when the subject matter tested is likely to be related to learning in the home (e.g., reading) rather than to schooling (e.g., mathematics) (Airasian & Madaus, 1983; Linn & Harnisch, 1981).

Schmidt (1983) identifies three major types of domains from which content to be covered can be drawn: *a priori* domains, curriculum-specific or learning-material-specific domains, and instructional material domains. Nitko (1983) suggests that "agents" not associated with local schools or particular programs tend to define *a priori* domains by using social criteria in judging what is important for all to learn. He goes on to suggest that test exercises in the National Assessment of Educational Progress (NAEP) as well as state assessment programs are examples of assessment instruments built from *a priori* domains since they specify content to be included without necessarily linking that content to specific instructional material or specific instructional events.

Cole & Nitko (1981) suggest that another design variable be considered in building tests to detect school and program effectiveness. They suggest that students require more time to acquire global skills and to grow in general educational development than to learn specific knowledges and skills. They suggest that tests measuring the former are less sensitive to measuring short term instructional efforts than tests measuring the latter.

Cooley (1977) and Leinhardt (1980) argue for the collection of relevant classroom variables and developing tests that are sensitive to differences between classrooms within-program. Leinhardt & Seewald (1981) describe several within-school, program, and classroom variables that are important to program evaluators and how to measure them. Mehrens and Phillips (Mehrens, 1984; Mehrens & Phillips, 1986; Phillips & Mehrens, 1988), however, found no significant differences on standardized tests from the use of different textbooks and different degrees of curriculum-test overlap when previous achievement and socioeconomic status were taken into account.

In the development of NELS:88 test items, efforts were made to take a middle road in the sense that our curriculum experts were instructed to select items that tapped general knowledge found in most curriculums but typically did not require a great deal of isolated factual knowledge. The emphasis was to be on understanding concepts and the measurement of problem-solving skills. However, it was thought necessary to assess the basic operational skills (e.g., simple arithmetic and algebraic operations) which are the foundations for successfully carrying out the problem-solving tasks.

The incorporation in the mathematics test of the relatively simple arithmetic and algebraic items which measure procedural or factual knowledges served two purposes. First, this subset of items provided better assessment for those low scoring students who were just beginning to develop their "basic mathematical skills". Second, these items should be able to provide a limited amount of diagnostic information about why some students are not able to successfully carry out the tasks defined in the typically more demanding problem-solving items. For example, students who are not proficient on the problem-solving items can be further divided into two groups based on their performance on the arithmetical/algebraic procedural skill items. One subgroup could not very well be proficient on the problem-solving items since they did not demonstrate sufficient skills on the simple arithmetical/algebraic procedures that are a necessary but not a sufficient condition for successful performance on the problem-solving tasks. The remaining subgroup, however, had sufficient grounding in the basics as demonstrated by their successful performance on the procedural items but were unable to carry out the logical operations necessary to complete the solutions to the problem solving items.

This hierarchical nature of the required skills is put to formal use in the development of behaviorally anchored proficiency level scales for reading, science and mathematics. This criterion-referenced interpretation is discussed further in the chapter describing the estimated scores.

This concern with respect to the maximal overlap doctrine is particularly relevant to the measurement of change over relatively long periods of exposure to varied educational treatments. That is, the two-year gaps between re-testings coupled with a very heterogeneous student population are quite likely to coincide with considerable variability in course taking experiences. This fact, along with the constraints on testing time, makes coverage of specific curriculum related knowledges very difficult. Also, as indicated above, specificity in the knowledges being tapped by the cognitive tests could lead to distortions in the gain scores due to forgetting of specific details. The impact on gain scores due to forgetting should be minimized if the cognitive battery increasingly emphasizes general concepts and development of problem solving abilities. This emphasis should increase as one goes to the tenth and twelfth grades. Students who take more high level courses, regardless of the specific course content, are likely to increase their conceptual understanding as well as gain additional practice in problem-solving skills.

At best any nationally based longitudinal achievement testing program must be a compromise that attempts to balance testing time burdens, the natural tensions between local curriculum emphasis and more general mastery objectives, and the psychometric constraints (in the NELS:88 case) in carrying out both vertical equating (year-to-year) and cross-sectional equating (form-to-form within year). NELS:88 fortunately did have the luxury of being able to gather cross-sectional pre-test data on the item pools. Thus we have been able to take into consideration not only the general curriculum relevance but whether or not the items demonstrate reasonable growth curves, as well as meet the usual item analysis parameter requirements for item quality.

The following sections contain descriptions of the content and format of each of the four achievement tests along with selected classical item statistics.

Reading

The reading test forms consisted of four or five reading passages, ranging in length from a single paragraph to a half-page. There are two forms of the reading test, differing in difficulty, in both the tenth and twelfth grade. Each passage in the reading tests (or forms) was followed by three to five multiple-choice questions addressing the students' ability to reproduce details of the text, translate verbal statements into concepts (comprehension), or draw conclusions based on the material presented (inference/evaluation). A total of 21 questions was presented in 21 minutes. The amount of time allowed for each question, which is relatively long compared to the other three content areas, takes into account the length of time needed for reading the passages before answering the questions.

The reading tests typically began with the least difficult passage followed by four or five relatively easy questions. The content/process specifications of the pool of items that made up NELS:88 reading forms across all grades and forms within grade are presented in Table 2.1. The percent answering each item correctly (P+) and the item-total correlations (biserials) are presented by grade, and by form within grade for the total population in Tables 2.2 and 2.3. The IRT parameters for the reading test are presented in appendix E-1. The P+ values and biserials are presented for those forms and grades for which they were administered. The more difficult items that differentiated the twelfth grade "high" form from the easier forms required comprehension of social studies material or inferences based on science material.

Appendices A-1 to A-5 present the P+'s and biserials for gender and racial/ethnic groups also. Tables 2.2 and 2.3 not only present the P+'s and biserials by form, but the reader can quickly identify the linking items for each of the forms. The linking items provide the overlap between forms that is necessary to put all scores on the same vertical scale, regardless of the form given. In general, we have tried to be conservative in the sense that we have more overlapping items than one typically finds in a vertically equated test battery.

Table 2.1
NELS:88 Reading Specifications
Content by Process by Test Forms[a]

Process	Content Area		
	Literary	Science	Social Studies/Other
Reproduction of Detail			
Test Form	3	1	-
8th Grade	3	1	-
10th Grade Low	2	1	1
10th Grade High	3	1	1
12th Grade Low	-	-	1
12th Grade High			
Comprehension of Thought			
Test Form	1	1	1
8th Grade	1	1	1
10th Grade Low	3	1	2
10th Grade High	-	2	4
12th Grade Low	-	1	8
12th Grade High			
Inferences and/or Evaluative Judgements			
Test Form	10	1	3
8th Grade	10	1	3
10th Grade Low	9	1	1
10th Grade High	6	1	3
12th Grade Low	4	3	3
12th Grade High			

[a]Entries in table are the number of items

Table 2.2
Reading: Proportion Correct

Item No.	Base Year	First Follow-up		Second Follow-up	
		Low	High	Low	High
Item 1	.95	.92		.93	
Item 2	.85	.80		.82	
Item 3	.82	.77		.80	
Item 4	.57	.50		.57	
Item 5	.55	.46		.56	
Item 6			.63		
Item 7			.55		
Item 8			.55		
Item 9			.66		
Item 10			.57		
Item 11			.84		
Item 12			.60		
Item 13			.76		
Item 14				.25	
Item 15	.60	.54	.86	.58	
Item 16	.41	.33	.67	.36	
Item 17	.49	.44	.81	.45	
Item 18	.61	.54			
Item 19	.39	.36	.52	.36	.57
Item 20	.59		.76		

Table 2.2
Reading: Proportion Correct (cont'd)

Item No.	Base Year	First Follow-up		Second Follow-up	
		Low	High	Low	High
Item 21		.65			
Item 22	.71	.62	.91	.63	.94
Item 23	.50	.48	.79	.53	.86
Item 24	.48	.41	.82	.47	.89
Item 25					.47
Item 26					.70
Item 27					.90
Item 28					.87
Item 29			.51		
Item 30			.63		
Item 31			.78		
Item 32			.45		
Item 33			.36		
Item 34					.59
Item 35					.32
Item 36					.50
Item 37					.42
Item 38	.46	.38		.48	
Item 39	.76	.71		.79	
Item 40	.54	.40			

Table 2.2
Reading: Proportion Correct (cont'd)

Item No.	Base Year	First Follow-up		Second Follow-up	
		Low	High	Low	High
Item 41	.54	.46		.54	
Item 42	.63	.55			
Item 43	.70	.67			
Item 44	.62	.55			
Item 45				.64	.84
Item 46				.42	.61
Item 47				.68	
Item 48				.35	.52
Item 49				.34	.56
Item 50					.77
Item 51					.49
Item 52					.43
Item 53					.44
Item 54					.30
Mean	.61	.55	.67	.55	.62
S.D.	.14	.15	.15	.18	.20
Unwtd	23643	9115	8717	7076	7154
Wtd N	2897540	1511539	1368601	1222645	1058046

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 2.3
Reading: R-Biserial

Item No.	Base Year	First Follow-up		Second Follow-up	
		Low	High	Low	High
Item 1	.60	.63		.64	
Item 2	.63	.61		.66	
Item 3	.65	.65		.67	
Item 4	.67	.59		.64	
Item 5	.67	.58		.62	
Item 6			.51		
Item 7			.53		
Item 8			.57		
Item 9			.70		
Item 10			.53		
Item 11			.72		
Item 12			.62		
Item 13			.70		
Item 14				.47	
Item 15	.65	.61	.68	.70	
Item 16	.63	.51	.61	.61	
Item 17	.68	.61	.69	.62	
Item 18	.57	.45			
Item 19	.44	.41	.41	.37	.43
Item 20	.64		.59		

Table 2.3
Reading: R-Biserial (cont'd)

Item No.	Base Year	First Follow-up		Second Follow-up	
		Low	High	Low	High
Item 21		.59			
Item 22	.75	.69	.75	.69	.66
Item 23	.55	.48	.66	.52	.61
Item 24	.65	.58	.73	.62	.65
Item 25					.46
Item 26					.47
Item 27					.45
Item 28					.62
Item 29			.50		
Item 30			.47		
Item 31			.65		
Item 32			.48		
Item 33			.41		
Item 34					.51
Item 35					.47
Item 36					.59
Item 37					.55
Item 38	.70	.61		.66	
Item 39	.74	.72		.69	
Item 40	.66	.52			

Table 2.3
Reading: R-Biserial (cont'd)

Item No.	Base Year	First Follow-up		Second Follow-up	
		Low	High	Low	High
Item 41	.53	.47		.50	
Item 42	.67	.64			
Item 43	.64	.58			
Item 44	.62	.53			
Item 45				.53	.66
Item 46				.33	.61
Item 47				.59	
Item 48				.45	.54
Item 49				.39	.60
Item 50					.60
Item 51					.47
Item 52					.47
Item 53					.44
Item 54					.45
Mean	.63	.57	.60	.57	.54
S.D.	.07	.08	.10	.11	.08

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Mathematics

Tables 2.4, 2.5 and 2.6 present the content by process specifications and the P+'s and biserials for the seven mathematics forms respectively. Appendices B-1 to B-7 give the P+'s and biserials for the gender and racial/ethnic groups. Appendix E-2 presents the IRT item parameters for the mathematics test. The biserials do drop below the desirable .45 - .50 range for some of the forms, primarily due to the restriction in range of abilities that occurs within a form. Inspection of Table 2.4 indicates that what distinguishes the "high" tenth and twelfth grade forms from the other forms is the increased emphasis on

Table 2.4
NELS:88 Math Specifications
Content by Process by Test Forms[a]

Process	Arithmetic	Algebra	Geometry	Data/Prob	Adv Topic
Skill/Knowledge					
Test Form	10	5	1	1	-
8th Grade	12	4	2	-	-
10th Grade Low	9	3	-	1	1
10th Grade Med	6	3	-	2	2
10th Grade High	10	4	2	-	-
12th Grade Low	7	2	-	1	1
12th Grade Med	1	2	-	1	2
12th Grade High					
Under/Comprehend					
Test Form	6	7	3	3	-
8th Grade	7	6	3	2	-
10th Grade Low	6	6	3	2	-
10th Grade Med	3	7	2	3	2
10th Grade High	6	5	3	3	-
12th Grade Low	4	6	4	2	-
12th Grade Med	1	5	7	1	3
12th Grade High					
Problem Solving					
Test Form	3	-	-	-	1
8th Grade	3	-	-	-	1
10th Grade Low	3	2	2	-	2
10th Grade Med	2	2	3	-	2
10th Grade High	4	-	2	-	1
12th Grade Low	4	3	5	-	1
12th Grade Med	2	4	9	1	1
12th Grade High					

[a]Entries in table are the number of items

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 2.5
Math: Proportion Correct

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 1	.56	.42	.67	.92	.52	.76	
Item 2					.46		
Item 3	.69	.50		.93	.58		
Item 4		.83			.90		
Item 5	.52	.37	.62	.90			
Item 6	.59	.45	.75		.58		
Item 7	.65	.47			.57		
Item 8	.51	.44	.71	.94	.44		
Item 9	.62	.49	.72	.95	.48	.78	
Item 10	.66	.51					
Item 11	.51	.37	.70	.96	.42	.78	
Item 12	.49	.35	.62	.93	.40	.74	
Item 13	.44	.31	.53	.87	.35		
Item 14		.71			.80		
Item 15	.41		.49	.88			
Item 16	.44	.26	.56	.84			
Item 17	.50		.56	.84			
Item 18	.47		.47	.79			
Item 19		.27					
Item 20		.27					
Item 21		.54			.51		
Item 22	.52	.30	.62	.90	.31	.73	
Item 23	.41	.27	.49	.87	.37	.60	
Item 24	.45		.49	.83		.53	.90
Item 25	.37		.41	.73		.46	.82

Table 2.5
Math: Proportion Correct (cont'd)

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 26	.35	.21	.49	.84	.22	.56	.86
Item 27							.40
Item 28	.50	.27	.58	.92	.31	.66	
Item 29	.71	.57			.96	.56	
Item 30	.79	.68	.82		.75	.86	
Item 31	.70	.63	.75		.66	.77	
Item 32	.52	.31	.59	.93	.35	.69	
Item 33	.79	.73	.88		.74	.90	
Item 34	.46		.49	.71	.43	.58	
Item 35	.59	.45	.69	.88	.43	.75	
Item 36	.52	.39	.58	.85	.41	.64	.89
Item 37	.38	.17	.46	.92	.20	.50	.95
Item 38	.45		.59	.92			
Item 39	.27	.31	.62	.92	.34	.72	.97
Item 40	.41	.32	.39	.66		.39	.80
Item 41						.27	.48
Item 42							.51
Item 43				.31		.20	.41
Item 44	.40	.23	.49	.86	.26	.58	.92
Item 45					.25	.31	.53
Item 46				.55			.71
Item 47				.45			.59
Item 48							.46
Item 49				.66			.90
Item 50	.56	.46	.61	.86	.44	.67	

Table 2.5
Math: Proportion Correct (cont'd)

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 51			.42	.77		.56	.91
Item 52				.53			.76
Item 53			.55	.83			
Item 54			.35	.69		.36	.81
Item 55			.34	.68		.36	.76
Item 56			.29	.60		.33	.71
Item 57			.29	.64		.36	.79
Item 58						.06	.15
Item 59						.15	.24
Item 60	.71	.54	.78		.65	.91	
Item 61	.79	.76	.91		.85	.93	
Item 62	.68	.55		.66			
Item 63	.65	.56	.73		.59	.73	
Item 64	.61	.33			.32		
Item 65		.23					
Item 66		.68			.80		
Item 67					.60		.93
Item 68					.14		.89
Item 69					.28	.40	.67
Item 70					.22	.45	.84
Item 71						.46	.59
Item 72						.33	.57
Item 73						.23	.57
Item 74							.41
Item 75							.54

Table 2.5
Math: Proportion Correct (cont'd)

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 76							.41
Item 77							.37
Item 78							.16
Item 79							.30
Item 80							.23
Item 81							.26
Mean	.54	.44	.58	.80	.48	.55	.62
S.D.	.13	.17	.15	.15	.19	.22	.24
Unwtd	23648	3199	9780	4814	2554	7717	3965
Wtd N	2897116	545728	1635418	689739	429799	1293720	557388

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 2.6
Math: R-Biserial

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 1	.60	.41	.51	.56	.42	.54	
Item 2					.45		
Item 3	.56	.31		.52	.40		
Item 4		.49			.53		
Item 5	.66	.44	.56	.55			
Item 6	.68	.49	.61		.48		
Item 7	.65	.45			.48		
Item 8	.60	.46	.63	.66	.43		
Item 9	.60	.40	.59	.68	.47	.61	
Item 10	.55	.38					
Item 11	.65	.48	.70	.93	.50	.72	
Item 12	.65	.41	.62	.75	.50	.65	
Item 13	.51	.40	.53	.56	.31		
Item 14		.51			.46		
Item 15	.69		.63	.58			
Item 16	.66	.43	.61	.54			
Item 17			.52	.45			
Item 18	.27		.26	.37			
Item 19		.36					
Item 20		.37					
Item 21		.40			.43		
Item 22	.70	.49	.61	.60	.44	.55	
Item 23	.60	.40	.54	.58	.38	.60	
Item 24	.45		.45	.52		.54	.50
Item 25	.58		.49	.53		.49	.40

Table 2.6
Math: R-Biserial (cont'd)

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 26	.54	.28	.60	.58	.32	.57	.37
Item 27							.55
Item 28	.69	.41	.62	.70	.50	.63	
Item 29	.51	.41		.73	.37		
Item 30	.50	.46	.46		.23	.36	
Item 31	.46	.31	.39		.33	.43	
Item 32	.64	.36	.61	.76	.44	.62	
Item 33	.59	.50	.61		.35	.44	
Item 34	.31		.23	.41	.21	.37	
Item 35	.57	.40	.47	.41	.34	.45	
Item 36	.54	.40	.46	.52	.37	.48	46
Item 37	.70	.33	.65	.65	.36	.64	.43
Item 38	.70		.60	.56			
Item 39	.62	.56	.65	.62	.55	.71	.41
Item 40	.32	.16	.30	.55		.37	.63
Item 41						.20	.49
Item 42							.48
Item 43				.38		.33	.40
Item 44	.63	.37	.51	.55	.41	.61	.51
Item 45					.16	.34	.38
Item 46				.52			.55
Item 47				.35			.37
Item 48							.58
Item 49				.59			.68
Item 50	.50	.31	.43	.49	.35	.46	

Table 2.6
Math: R-Biserial (cont'd)

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 51			.49	.55		.61	.58
Item 52				.62			.65
Item 53			.53	.51			
Item 54			.35	.67		.49	.57
Item 55			.40	.56		.45	.58
Item 56			.34	.48		.42	.44
Item 57			.49	.53		.53	.51
Item 58						.25	.56
Item 59						.17	.48
Item 60	.69	.56	.66		.65	.79	
Item 61	.51	.57	.63		.58	.59	
Item 62	.71	.49			.50		
Item 63	.45	.41	.29		.44	.30	
Item 64	.76	.55			.50		
Item 65		.28					
Item 66		.47			.45		
Item 67					.43		.44
Item 68					.37		.61
Item 69					.38	.39	.45
Item 70					.28	.60	.51
Item 71						.22	.35
Item 72						.25	.48
Item 73						.52	.59
Item 74							.40
Item 75							.54

Table 2.6
Math: R-Biserial (cont'd)

Item No.	Base Year	First Follow-up			Second Follow-up		
		Low	Mid	High	Low	Mid	High
Item 76							.65
Item 77							.61
Item 78							.43
Item 79							.44
Item 80							.64
Item 81							.59
Mean	.58	.42	.52	.57	.41	.48	.51
S.D.	.11	.09	.12	.11	.10	.15	.09

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

understanding concepts and problem solving in the areas of geometry, data/probability, and advanced topics. Advanced topics included pre-calculus items and/or analytic geometry items. It should be kept in mind that while an item may be classified as a geometry item, it more often than not requires both algebraic and numeric skills for a correct solution. Similarly, the algebra items almost always require some facility in arithmetic to arrive at the correct solution. To the extent that any discipline tends to have a "building block" structure, the resulting assessment must also reflect the building block nature of the knowledge domain.

This hierarchical knowledge domain has its advantages and disadvantages. The advantage of a hierarchical knowledge domain is that it typically generates a large general factor which is a prerequisite for the item response theory (IRT) approach to the vertical scaling necessary for measuring longitudinal change on the same scale. One added benefit of the hierarchical knowledge domain is that it facilitates the interpretation of various ascending points along the vertical scale. That is, score points along the scale can be assigned a meaning to the extent they reflect different proficiency levels along the knowledge hierarchy. In this sense knowledge hierarchies allow one to have multiple criterion-referenced points along the vertical scale. The primary disadvantage is that subscores based on content areas are not likely to have much differential validity since virtually all mathematics items incorporate knowledges from many different content areas. In Chapter 4 on score estimation, more details will be presented on how both normative scores and mastery or proficiency score estimates were obtained in reading, science, and mathematics.

Science

Table 2.7 presents the content by process item specifications for the science forms.

Table 2.7
NELS:88 Science Specifications
Content by Process by Test Forms[a]

Process	Earth Sci	Chem	Sci Meth	Life Sci	Phy Sci
Skill/Knowledge					
Test Form					
8th Grade	5	2	-	3	-
10th Grade	3	2	-	2	1
12th Grade	3	3	-	3	1
Under/Comprehend					
Test Form					
8th Grade	2	2	1	2	-
10th Grade	2	1	1	2	1
12th Grade	1	-	3	1	-
Problem Solving					
Test Form					
8th Grade	1	3	2	2	-
10th Grade	-	3	1	3	2
12th Grade	-	3	1	2	4

[a] Entries in table are the number of items

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

The science tests were only grade level adaptive. That is, everyone within grade received the same form. The higher grade level forms (tenth and twelfth) were modified by adding more advanced material to minimize ceiling effects. Tables 2.8 and 2.9 present the P+'s and biserials for the items in each grade level form for the total population. Appendices C-1 to C-3 show the P+'s and biserials for gender and racial/ethnic groups. Appendix E-3 presents the IRT parameters for the science test.

Table 2.8
Science: Proportion Correct

Item No.	Base Year	First Follow-up	Second Follow-up
Item 1	.70		
Item 2	.79		
Item 3	.64	.72	
Item 4	.67	.74	.78
Item 5	.76	.78	.81
Item 6	.76	.84	.88
Item 7	.65		
Item 8	.57		
Item 9	.64		
Item 10	.53	.59	.65
Item 11	.48		
Item 12	.66	.73	.73
Item 13	.72		
Item 14	.53	.65	.70
Item 15	.39	.54	.56
Item 16	.46	.56	.58
Item 17	.42	.57	.63
Item 18	.45	.58	.65
Item 19	.42	.54	.59
Item 20	.41	.50	

Table 2.8
Science: Proportion Correct (cont'd)

Item No.	Base Year	First Follow-up	Second Follow-up
Item 21	.42	.51	
Item 22	.37	.46	.47
Item 23	.39	.50	
Item 24	.33	.42	.45
Item 25	.22	.32	
Item 26		.52	.61
Item 27		.28	.32
Item 28			.73
Item 29		.49	.58
Item 30		.50	.58
Item 31			.59
Item 32		.26	.34
Item 33		.56	.64
Item 34		.47	
Item 35			.43
Item 36			.43
Item 37			.29
Item 38			.13
Mean	.54	.55	.57
S.D.	.15	.14	.17
Unwtd	23616	17684	14134
Wtd N	2889974	2849102	2262896

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 2.9
Science: R-Biserial

Item No.	Base Year	First Follow-up	Second Follow-up
Item 1	.57		
Item 2	.51		
Item 3	.48	.53	
Item 4	.45	.51	.53
Item 5	.71	.71	.70
Item 6	.67	.70	.67
Item 7	.50		
Item 8	.46		
Item 9	.51		
Item 10	.53	.60	.65
Item 11	.41		
Item 12	.57	.61	.63
Item 13	.54		
Item 14	.65	.71	.73
Item 15	.47	.49	.47
Item 16	.42	.52	.54
Item 17	.49	.66	.71
Item 18	.54	.61	.61
Item 19	.50	.60	.62
Item 20	.35	.47	
Item 21	.39	.49	
Item 22	.38	.46	.46
Item 23	.27	.38	
Item 24	.56	.59	.62
Item 25	.37	.51	

Table 2.9
Science: R-Biserial (cont'd)

Item No.	Base Year	First Follow-up	Second Follow-up
Item 26		.60	.64
Item 27		.55	.65
Item 28			.52
Item 29		.63	.69
Item 30		.55	.60
Item 31			.50
Item 32		.56	.67
Item 33		.62	.65
Item 34		.44	
Item 35			.56
Item 36			.33
Item 37			.31
Item 38			.26
Mean	.49	.56	.57
S.D.	.10	.08	.12

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

History/Citizenship/Geography

Tables 2.10, 2.11 and 2.12 present the item content specifications, P+'s and biserials respectively.

Table 2.10
NELS:88 History Specifications Content by Test Forms

	Cit/Govt	Am Hist	Geog
8th Grade	13	14	3
10th Grade	8	19	3
12th Grade	12	15	3

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 2.11
History/Citizen/Geography: Proportion Correct

	Base Year	First Follow-up	Second Follow-up
Item 1	.69	.83	.89
Item 2	.49	.64	.66
Item 3		.63	
Item 4	.48	.56	
Item 5	.55	.68	.71
Item 6	.43	.50	.54
Item 7	.77	.83	
Item 8	.58	.67	.76
Item 9	.42	.52	.59
Item 10	.47	.52	.61
Item 11	.45	.44	.57
Item 12			.41
Item 13	.48	.53	.65
Item 14	.78	.80	
Item 15	.66	.72	.80
Item 16	.90	.91	
Item 17	.80	.85	
Item 18	.24	.28	.56
Item 19	.84	.91	.96
Item 20			.43
Item 21	.35	.44	.59
Item 22	.86		
Item 23	.84		
Item 24	.91		
Item 25	.88		

Table 2.11
History/Citizen/Geography: Proportion Correct (cont'd)

	Base Year	First Follow-up	Second Follow-up
Item 26	.91		
Item 27	.76	.80	.91
Item 28			.52
Item 29	.66	.74	
Item 30	.70	.81	
Item 31	.54	.67	.78
Item 32		.32	.43
Item 33	.47	.60	.72
Item 34	.59	.51	
Item 35		.71	
Item 36			.25
Item 37	.52	.56	.68
Item 38		.45	
Item 39		.42	
Item 40			.63
Item 41			.70
Item 42			.56
Item 43			.64
Item 44			.55
Item 45			.29
Item 46			.35
Item 47			.20
Mean	.63	.63	.60
S.D.	.19	.17	.18
Unwtd N	23525	17591	14063
Wtd N	2880468	2841095	2253399

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 2.12
History/Citizenship/Geography: R-Biserial

	Base Year	First Follow-up	Second Follow-up
Item 1	.63	.66	.67
Item 2	.53	.62	.68
Item 3		.40	
Item 4	.57	.67	
Item 5	.53	.58	.58
Item 6	.48	.59	.68
Item 7	.66	.72	
Item 8	.59	.67	.69
Item 9	.42	.46	.54
Item 10	.60	.63	.69
Item 11	.47	.49	.61
Item 12			.44
Item 13	.50	.52	.57
Item 14	.59	.62	
Item 15	.61	.61	.63
Item 16	.76	.78	
Item 17	.58	.64	
Item 18	.29	.46	.69
Item 19	.64	.68	.56
Item 20			.53
Item 21	.36	.59	.71
Item 22	.61		
Item 23	.49		
Item 24	.78		
Item 25	.67		

Table 2.12
History/Citizenship/Geography: R-Biserial (cont'd)

	Base Year	First Follow-up	Second Follow-up
Item 26	.79		
Item 27	.74	.77	.74
Item 28			.49
Item 29	.60	.69	
Item 30	.48	.58	
Item 31	.55	.60	.66
Item 32		.52	.55
Item 33	.48	.55	.60
Item 34	.64	.62	
Item 35		.46	
Item 36			.28
Item 37	.61	.65	.68
Item 38		.44	
Item 39		.31	
Item 40			.60
Item 41			.46
Item 42			.60
Item 43			.65
Item 44			.50
Item 45			.48
Item 46			.42
Item 47			.30
Mean	.58	.59	.58
S.D.	.11	.11	.11

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

There was no attempt to design process specifications into the H/C/G test. Appendices D-1 to D-3 show the P+'s and biserials for gender and racial/ethnic groups. Appendix E-4 presents the IRT parameters for the H/C/G test.

In summary, for almost all content areas the average P+'s for the grade level forms and the forms within grade are in the targeted middle ranges, i.e., .45 to .65. This is a desirable range because maximal discrimination in the sense of differentiation between people occurs at the P+ of .5. The one exception is the high level mathematics form in the tenth grade. The high level tenth grade mathematics form turned out to be easier than predicted from the field test statistics. This tendency for some potential ceiling effects in the high tenth grade mathematics form was somewhat reduced when all three time points were pooled and Bayesian IRT procedures applied which tend to "shrink" in both item parameters and scores within subpopulations. This Bayesian procedure will be discussed in more detail in the next section.

The biserials were pretty much on target yielding for the most part quite respectable averages, i.e., .50 or greater for most test forms. This is a desirable target since experience suggests that tests that achieve this average biserial level tend to approach test reliabilities in the middle eighties with as few as 20 items.

Chapter 3
IRT Scaling for Longitudinal Measurement and Equating to Earlier Cohorts

In order to accurately measure the extent of cognitive gains at both the group and individual level, the eighth grade tests and the various forms of the tenth and twelfth grade tests must be calibrated on the same scale. The most convenient way of doing this is to use Item Response Theory (IRT). In order to successfully carry out such a calibration, the eighth, tenth, and twelfth grade items should be relatively unifactorial within a subject area, say mathematics or reading, with the same dominant factor underlying all test forms. This suggests that there should be a common set of anchor items across adjacent forms and that most, but not necessarily all, content areas be represented in all grade forms. Increments in difficulty demanded in ascending grade forms (8, 10, 12) can be accomplished by: (1) increasing the problem-solving demands within the same familiar content areas and (2) including content in the later forms (in particular twelfth grade) that tap materials normally found in the advanced course sequence but build on skills learned earlier in the sequence.

As indicated earlier, Item Response Theory (IRT, see Lord, 1980) was used in calibrating the various forms within each content area. A brief background on IRT follows with additional information on the Bayesian approach taken here.

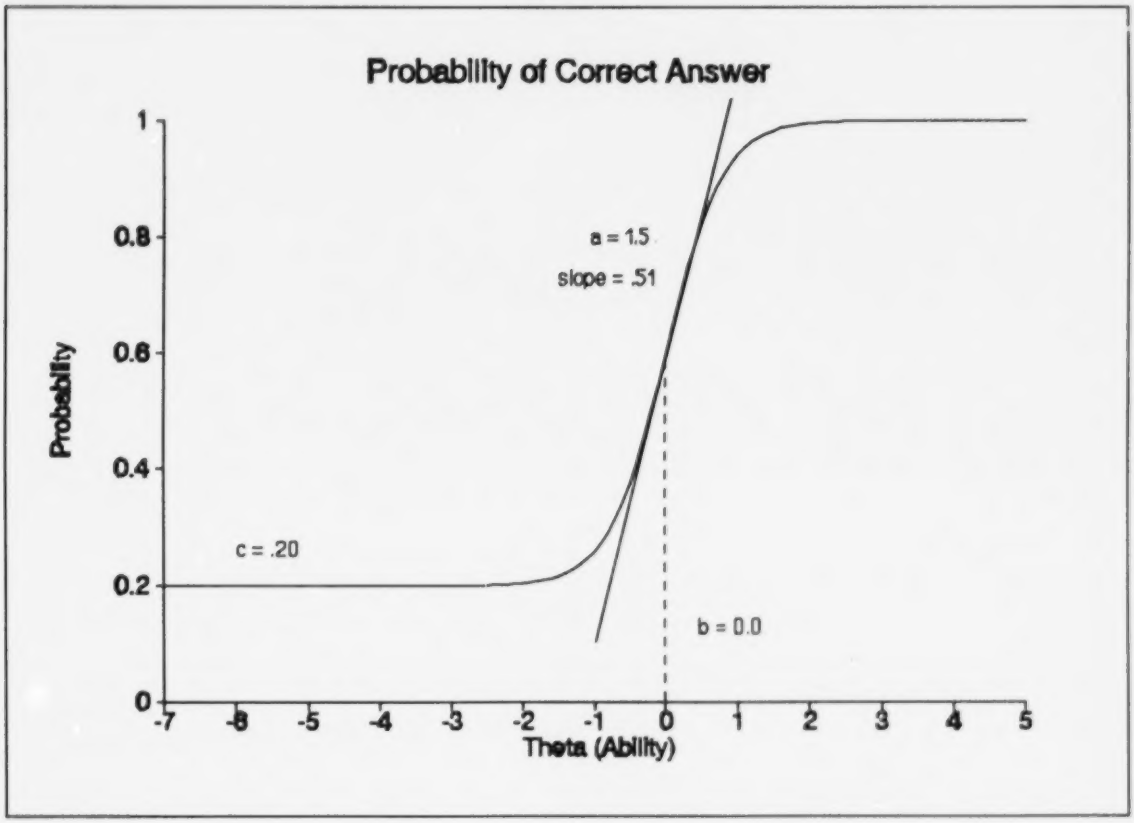
The underlying assumption of Item Response Theory (IRT) is that a test taker's probability of answering an item correctly is a function of his or her ability level for the construct being measured, and of one or more characteristics of the test item itself. The three-parameter IRT logistic model uses the pattern of right, wrong, and omitted responses to the items administered in a test form, and the difficulty, discriminating ability, and "guess-ability" of each item, to place each test taker at a particular point, θ (theta), on a continuous ability scale. Figure 3.1 shows a graph of the logistic function for a hypothetical test item. The horizontal axis represents the ability scale, theta. The point on the vertical probability axis corresponding to the height of the curve at a given value of theta is the estimated probability that a person of that ability level will answer the test item correctly. The shape of the curve is given by the following equation describing the probability of a correct answer on item i as:

$$P_i(\theta) = c_i + \frac{(1-c_i)}{1 + e^{-1.702 * a_i(\theta - b_i)}}$$

where θ = ability of the test taker
a_i = discrimination of item i, or how well the item distinguishes between ability levels at a particular point
b_i = difficulty of item i
c_i = "guessability" of item i

The "c" parameter represents the probability that a test taker with very low ability will answer the item correctly. In the graph above, 20% of test takers with a very low level of mastery of the test material guessed the correct answer to the question. The c parameter will not necessarily be equal to 1/(# options), e.g., .25 for a 4-choice item. Some response options may, for unknown reasons, be more attractive than random guessing, while others may be less likely to be chosen.

Figure 3.1

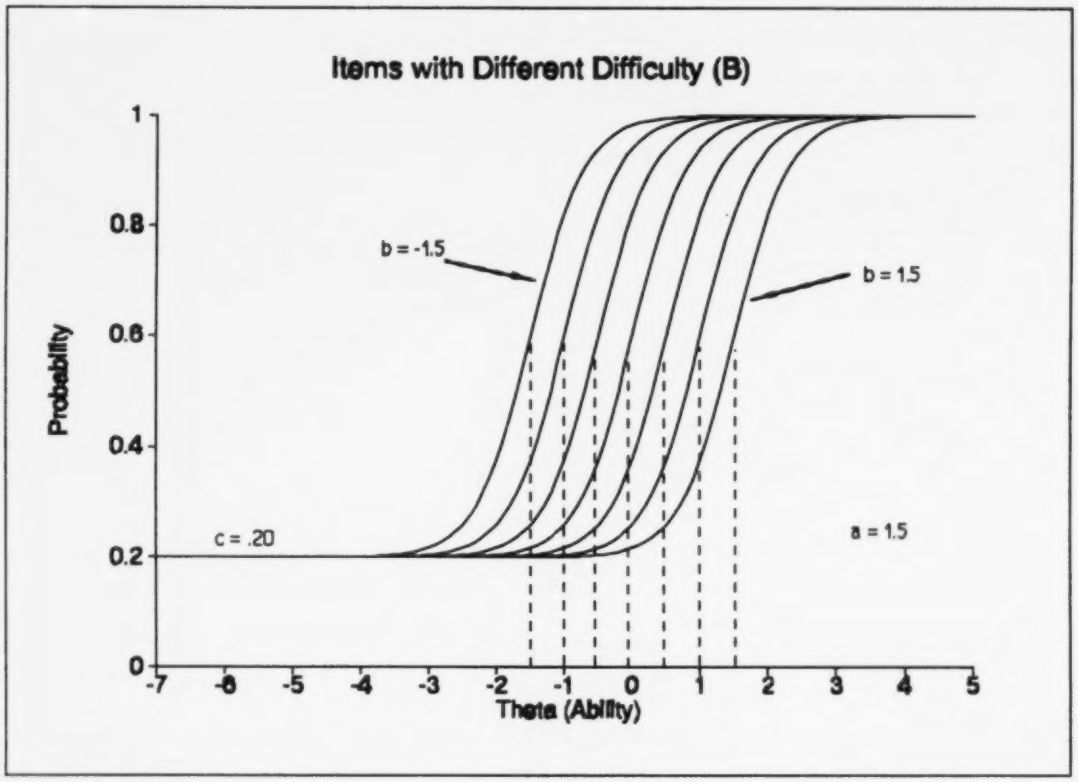


Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

The IRT "b" parameters correspond to the difficulty of the items, represented by the horizontal axis in the ability metric. In Figure 3.1, b = 0.0 means that test takers with $\theta = 0.0$ have a probability of getting the answer correct that is equal to halfway between the guessing parameter and 1. In this example, 60% of people at this ability level answered the question correctly. B also corresponds to the point of inflection of the logistic function. This point occurs farther to the right for more difficult items, and farther to the left for easier ones. Figure 3.2 is a graph of the logistic functions for seven different test items, all with the same "a" and "c" parameters, and with difficulties ranging from b = -1.5 to b = 1.5. For each of these hypothetical questions, 60% of test takers whose ability level matches the difficulty of the item are likely to answer correctly. Fewer than 60% will answer correctly at values of theta (ability) that are less than b, and more than 60% at $\theta > b$.

The discrimination parameter, "a", has perhaps the least intuitive interpretation of all. It is proportional to the slope of the logistic function at the point of inflection. Items with a steep slope are said to discriminate well. In other words, they do a good job of discriminating, or separating, people whose ability level is below the calibrated difficulty of the item (who are likely to get it right at only about the guessing rate) from those of ability higher than the item "b", who are nearly certain to answer correctly. By contrast, an item with a relatively flat slope is of little use in determining whether a person's

Figure 3.2

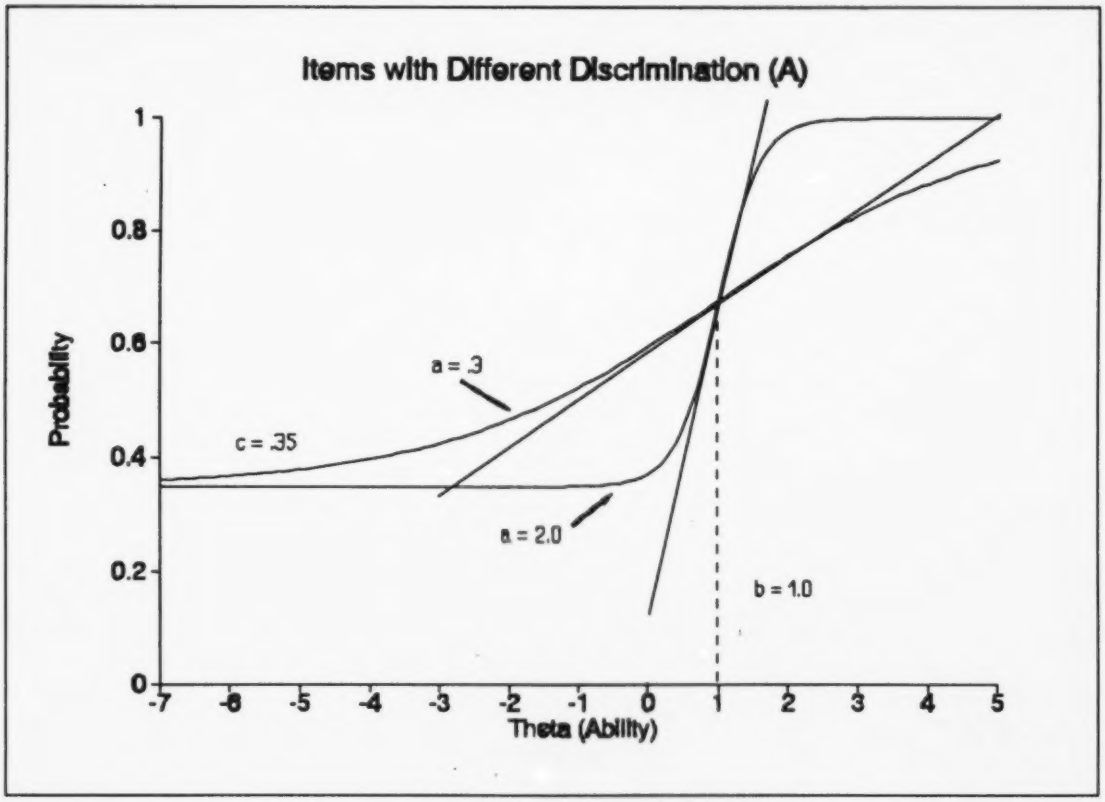


Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

correct placement along the continuum of ability is above or below the difficulty of the item. This idea is illustrated by Figure 3.3, representing the logistic functions for two test items having the same difficulty and guessing parameters, but different discrimination. The test item with the steeper slope (a = 2.0) provides useful information with respect to whether the test taker's ability level is above or below the difficulty level, 1.0, of the item: if the answer to this item was incorrect, the person very likely has an ability below 1.0; if the answer is correct, the test taker probably has a θ greater than 1.0, or guessed successfully. A series of many such highly discriminating items, with a range of difficulty levels (b parameters) such as those shown in Figure 3.2, will do a good job in narrowing the choice of probable ability level. Conversely, the flatter curve in Figure 3.3 represents a test item with a low discrimination parameter (a=.3). There is little difference in proportion of correct answers for test takers several points apart on the range of ability. So knowing whether a person's response to such an item is correct or not contributes relatively little to pinpointing his or her correct location on the horizontal ability axis.

BILOG or PARSCALE (Muraki & Bock, 1991) computer programs compute marginal maximum-likelihood estimates of IRT parameters that best fit the responses given by the test takers. The procedure calculates a, b, and c parameters for each test item, iterating until convergence within a specified level of accuracy is reached. Comparison of the IRT-estimated probability with the actual proportion of correct answers to a test item for examinees grouped by ability provides a means of evaluating the appropriateness

Figure 3.3



Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

of the model for the set of test data for which it is being used. A close match between the IRT-estimated curves and the actual data points means that the theoretical model accurately represents the empirical data.

Once a pool of test items exists whose parameters have been calibrated on the same scale as the test takers' ability estimates, a person's probability of a correct answer for each item in the pool can be computed, even for items that may not have been administered to that individual. The IRT-estimated number correct for any subset of items is simply the sum of the probabilities of correct answers for those items. Consequently, the score is typically not a whole number.

In addition to providing a mechanism for estimating scores on items that were not administered to every individual, IRT has advantages over raw number-right scoring in the treatment of guessed and omitted items. By using the overall pattern of right and wrong responses to estimate ability, it can compensate for the possibility of a low ability student guessing several hard items correctly. If answers on several easy items are wrong, a correct difficult item is, in effect, assumed to have been guessed. Omitted items are also less likely to cause distortion of scores, as long as enough items have been answered right and wrong to establish a clear pattern. Raw number-right scoring, in effect, treats omitted items as if they had been answered incorrectly. While this may be a reasonable assumption in a motivated

test, where it is in students' interest to try their best on all items, this may not always be the case in NELS:88.

As indicated earlier, a longitudinal growth study by its very nature consists of subpopulations defined by differing ability levels. That is, after all the assessments have been completed (three assessments in NELS:88) there are at least three recognizable subpopulations of different ability levels, which are tied to the time of testing. For example, the base year subpopulation will have, on average, a lower expected level of performance, than that found in each of the remaining two follow-ups. Similarly the average performance of the tenth graders will be lower than that of the twelfth graders. For those content areas in which multilevel adaptive testing was implemented, there are more than three definable ability level populations. In mathematics there were seven forms differing in difficulty, and thus there are seven ability groups which could be expected to differ in performance. In reading there were five forms, and thus the potential for having five subpopulations with differing levels of performance.

In the past, when LOGIST (Wingersky, Barton & Lord, 1982) was the only reliable and documented three parameter computer program applicable in this area, one psychometrically acceptable procedure for vertical scaling in a longitudinal study would be to estimate the base year item parameters and fix their values at their base year quantities. When the first follow-up becomes available, item parameters would be estimated for only those items unique to the first follow-up. The scale is anchored by the items that were common to both the base year and the first follow-up, and which had their values fixed at their base year quantities. Variations that are improvements on this approach might include pooling the two waves of data and re-estimating all item parameters using all the available data and then using common item equating approaches such as the Stocking & Lord (1983) transformation to find linking constants that optimally match proportion correct on the item pool conditional on the scale (ability) scores. This second approach uses all the data in estimating the item parameters and thus could be expected to yield more stable item parameter estimates. The pooling of all time points and re-estimating the item parameters, of course can lead to a re-making of history in a longitudinal study where intermediate reports are published before all the data from all the time periods is available. That is, eighth grade scores that have been reported and analyzed might later be modified when the tenth and twelfth grade data became available. The use of all data points over time, however, is the preferable method because it is the one method which can provide stable estimates of both the item traces and latent trait scores throughout the entire ability distribution. This procedure was used in the vertical equating that was carried out for the High School and Beyond (Rock et al., 1985; Rock & Pollack, 1987).

The major problem with the above LOGIST approaches is that there is no easy way to incorporate into the item parameters and latent trait score estimation procedure prior knowledge about what ability distribution an individual comes from. This shortcoming is particularly crucial in its impact on measuring change in longitudinal studies. The inability of LOGIST and/or other non-Bayesian approaches to IRT is that they have no acceptable way of coping with "perfect" i.e., all correct scores. For example, some very advanced individuals who took the high level mathematics form in grade ten got all the items correct. In conditional maximum likelihood approaches such as LOGIST, such scores are undefined or are given some arbitrary high value. Yet we know these individuals, while gifted, probably will not get perfect scores when they eventually take the high level twelfth grade form. Does this mean that they are less knowledgeable in grade 12 than in grade 10? Probably not. In fact almost nobody got all the items correct in the "hardest" form in twelfth grade. Thus if they had been given the hard items from the twelfth grade "high" form when they were tenth graders they would indeed have had less than perfect scores, and if the same set of items were repeated they would more than likely show gains.

Pooling all three time points, which amounts to pooling all the items as well as people (in a sense pooling all available information) and recomputing all the item parameters using Bayesian priors reflecting the ability distributions associated with each particular test form, provides for an empirically based shrinkage to more reasonable item parameters and ability scores (Muraki & Bock, 1991). The fact that the total item pool is used in conjunction with the Bayesian priors leads to shrinking back the extreme item parameters as well as the perfect scores to a more reasonable quantity, which in turn allows for the potential of some gains even in the uppermost tail of the distribution. Each of the test forms (the eighth, tenth and twelfth grade forms, and in the case of reading and math, the multiple forms within year) is treated as a separate subpopulation with its own ability distribution. The amount of shrinkage is a function of the distance from the subgroup means and the relative reliability of the score being estimated. Theoretically this approach has much to recommend it. In practice, it has to have reasonable estimates of the difference in ability levels among the subpopulations in order to incorporate realistic priors. Essentially, the scales are determined by the linking items, and the initial prior means for the subgroups are in turn determined by the differential performance of the subpopulations on these linking items. For this reason we have designed the item pool to have an overabundance of items linking forms. This approach, using adaptive testing procedures combined with Bayesian procedures that allow for priors on both ability distributions and on the item parameters, is needed in longitudinal studies to minimize ceiling and floor effects.

A multiple group version of the PARSCALE computer program (Muraki & Bock, 1991) that was developed for NAEP allows for both group ability priors and item priors. A publicly available multiple group version of the BILOG (Mislevy & Bock, 1982) computer program called BIMAIN (Muraki & Bock, 1987, 1991) has many of the same capabilities for dichotomously scored items only. Since the PARSCALE program was applied to dichotomously scored items in the NELS:88 vertical scaling, its estimation procedure is identical to the multiple group version of BILOG or BIMAIN. PARSCALE uses a marginal maximum likelihood estimation approach and thus does not estimate the individual ability scores when estimating the items parameters but assumes that the ability **distribution** is known for each subgroup. Thus the posterior distribution of item parameters is proportional to the product of the likelihood of observing the item response vector, based on the data and conditional of the item parameters and subgroup membership, and the assumed prior ability distribution for that subgroup. More formally, the general model in terms of item estimation is the same as that used in NAEP and described in some detail by Yamamoto & Mazzeo (1992; p. 158) as follows:

$$\begin{aligned} L(\beta) &= \Pi_g \Pi_{j:g} \int_0 P(x_{j:g}|\theta,\beta) f_g(\theta) d(\theta) \\ &\sim \Pi_g \Pi_{j:g} \Sigma_k P(x_{j:g}|\theta = X_k,\beta) A_g(X_k). \end{aligned} \qquad (1)$$

In equation (1), $P(x_{j:g}|\theta,\beta)$ is the conditional probability of observing a response vector $x_{j:g}$ of person j from group g, given proficiency θ and vector of item parameters $\beta = (a_1,b_1,c_1,\ldots,a_p,b_p,c_p)$, and $f_g(\theta)$ is a population density for θ in group g. Prior distributions on item parameters can be specified and used to obtain Bayes modal estimates of these parameters (Mislevy, 1984). The proficiency densities can be assumed known and held fixed during item parameter estimation or can be estimated concurrently with item parameters.

The $f_g(\theta)$ in (1) are approximated by multinomial distributions over a finite number of quadrature points, where X_k for $k = 1,...,q$, denotes the set of points and $A_g(X_k)$ are the multinomial probabilities at the corresponding points that approximate $f_g(\theta)$ at $\theta = X_k$. If the data are from a single population with an assumed normal distribution, Gauss-Hermite quadrature procedures provide an optimal set of points and weights to best approximate the integral in (1) for a broad class of smooth functions. For more general f or for data from multiple populations with known densities, other sets of points (e.g., equally spaced points) can be substituted, and the values of $A_g(X_k)$ may be chosen to be the normalized density at point X_k (i.e., $A_g(X_k) = f_g(X_k)/\Sigma_k f_g(X_k)$).

Maximization of $L(\beta)$ is carried out by an application of an EM algorithm (Dempster, Laird & Rubin, 1977). When population densities are assumed known and held constant during estimation, the algorithm proceeds as follows. In the E step, provisional estimates of item parameters and the assumed multinomial probabilities are used to estimate expected sample sizes at each quadrature point for each group (denoted $\hat{N}_{gk}$), as well as over all groups (denoted $\hat{N}_k = \Sigma_g \hat{N}_{gk}$). These same provisional estimates are also used to estimate an expected frequency of correct responses at each quadrature point for each group (denoted $\hat{r}_{gik}$), and over all groups (denoted $\hat{r}_{ik} = \Sigma_g \hat{r}_{gik}$). In the M step, improved estimates of the item parameters are obtained by treating the $\hat{N}_{gk}$ and $\hat{r}_{ik}$ as known and carrying out maximum likelihood logistics regression analysis to estimate the item parameters β, subject to any constraints associated with prior distributions specified for β.

The user of the multiple group version of PARSCALE has the option of fixing the priors on the ability distribution or allowing the posterior estimate to update the previous prior and combine with the data-based likelihood to arrive at a new set of posterior estimates after each major EM cycle. If one wishes to update on each cycle, one can continue to constrain the priors to be normal or their shape can be allowed to vary. The NELS:88 approach was to allow for updating the prior but with the normality assumption. It was our experience that the "smoothing" that came from the updated normal priors led to less "jagged" looking ability score distributions and did not tend to overfit the item parameters. It has been our experience that lack of fit in the item parameter distribution would simply be absorbed in the shape of the ability distribution if the updated ability distribution were allowed to take any shape. A similar procedure was used in estimating the item parameters in the National Adult Literacy Study (NALS) (Kirsch et al. 1993).

Appendices E-1 to E-4 present the final item parameters for each of the content areas. The location of each item within each test form is also given, as well as the number of possible answer choices for each. Table 3.1 summarizes the means, standard deviations and ranges of the item parameters by content areas.

Table 3.1
Means, Standard Deviations and Ranges of IRT Parameters

	Number of Items	Mean	S.D.	Low	High
Reading					
A	54	0.9052	0.2901	0.3219	1.7607
B	54	0.0755	1.0757	-2.5174	2.3409
C	54	0.1494	0.1135	0.0000	0.4523
Math					
A	81	0.9529	0.3119	0.4168	2.1455
B	81	0.2987	1.4750	-2.9487	3.2030
C	81	0.1558	0.1091	0.0000	0.4388
Science					
A	38	0.8778	0.3186	0.3269	1.5459
B	38	0.0387	1.0006	-1.9340	2.4048
C	38	0.1850	0.1280	0.0000	0.3886
History					
A	47	1.0812	0.3802	0.2955	2.0344
B	47	-0.1899	1.2413	-2.6938	2.2582
C	47	0.2187	0.1286	0.0000	0.5162

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

With respect to interpreting the item parameters, "a" parameters (the discrimination parameter) should each be over .50. "a" parameters in the neighborhood of 1.0 or above are considered very good. As described earlier, the a parameter indicates the usefulness of the item in discriminating between points on the ability scale. The b parameter, item difficulty, should span the range of abilities being measured. Item difficulties should be concentrated in the range of abilities that contains most of the test takers. Test items provide the most information when their difficulty is close to the ability level of the examinees. Items that are too easy or too difficult for most of the test takers are of little use in discriminating between them. Ideally the "c" parameter (the probability of a low ability person guessing correctly) should be less than .25 for four choice items, but they may vary with difficulty, and of course the number of options. Most content areas had a mixture of four choice and five choice items. The H/C/G test had some two

choice items, and thus the somewhat elevated guessing parameters. In general, the item parameters meet these standards.

It should be remembered that the solution to equation 1 above finds those item parameters that maximize the likelihood across all groups (forms): seven in mathematics, five in reading, and three each in science and H/C/G. The present version of the multiple group PARSCALE only saves the subpopulation means and standard deviations and not the individual expected *a posteriori* (EAP) scores. The individual EAP scores which are the means of the posterior distributions of the latent variate, were obtained from the bgroup conditioning program which uses the Gaussian quadrature procedure. This variation is virtually equivalent to conditioning (e.g., see Mislevy, et al. 1992) on a set of "dummy" variables defining which ability subpopulation an individual comes from. The one difference is that the group variances are not restricted to be equal as in the standard conditioning procedure.

In summary, equation one finds the item parameters that maximize the likelihood function across all groups (forms and grades) simultaneously. The items can be put on the same vertical scale because of the linking items that are common to either adjacent forms or some subset of forms. Using the performance on the common items the subgroup means can be located along the vertical scale. Since marginal maximum likelihood estimation requires only an assumed ability density function in the estimation of item parameters, individual ability scores are not estimated in the item parameter estimation step, only the subgroup means and variances are estimated. The bgroup program then estimates the individual ability scores as the mean of an individual's posterior distribution. The posterior distributions for each individual at any given step in the bgroup iteration are the product of the likelihood of observing that pattern of "0"'s and "1"'s in the item response vector conditional on the item parameters and subgroup membership and the prior ability distribution. The prior ability distributions are assumed normal with a mean and variance from their subgroup. At each succeeding step in the iterations the previous posterior distribution becomes the new prior until the iterations converge.

Conditional independence is an assumption of all IRT models, but as Mislevy, et al., (1992) point out, not likely to be generally true. However, if one thinks of IRT-based scores as a summarization of essentially the largest latent factor underlying a given item pool, then small violations are of little significance. To insure that there were no substantive violations of this assumption, factor analyses were carried out on the grade 8 forms to insure a large dominant factor underlying each content area. These results were reported by Rock & Pollack (1987). Since students in the tenth and twelfth grade took different forms, factor analysis was no longer appropriate. However, all item traces were inspected to insure a good fit throughout the ability range. More importantly, estimated proportions correct by item by grade were also estimated in order to insure that the IRT model was both reproducing the item P+'s and there was no particular bias in favor of any particular grade. Since the item parameters were estimated using a model that maximizes the goodness-of-fit across the subpopulations, including grades, one would not expect much difference here. When the differences were summed across all items for each test, the maximum discrepancy between observed and estimated proportion correct for the whole test was .7 of a scale score point for grade twelve mathematics whose score scale had a range of 0 to 81. The IRT estimates tended to slightly underestimate the observed proportions. However, no systematic bias was found for any particular grade. Appendices F-1 to F-4 provide discrepancies by item as well as for totals aggregated across all items.

Differential Item Functioning (DIF)

Differential Item Functioning (DIF) as defined here attempts to identify those items showing an unexpectedly large difference in item performance between a focal group (e.g. Black students) and a

reference group (e.g. White students) when the two groups are "blocked" or matched on their total score. It should be noted that any such strictly internal analysis, i.e., without an external criterion, cannot detect bias when that bias pervades all items in the test (Cole & Moss, 1989). It can only detect differences in the relationships among items that are anomalous in some group in relation to other items. In addition such approaches can only identify the items where there is unexpected differential **performance**, they cannot directly imply **bias**. A determination of bias implies not only that differential performance on the item is related to subgroup membership, but also that the difference is **unfairly** associated with subgroup membership. That is, the difference is due to an attribute **not related** to the construct being measured. As Cole & Moss (1989) point out, items so identified must still be interpreted in light of the intended meaning of the test scores before any conclusion of bias can be drawn. It is not entirely clear how the term item bias applies to academic achievement measures given to students with different patterns of exposure to content areas. For example, some students may take more algebra after eighth grade while another group may take less algebra and more geometry. Both groups may have similar total scores but for one group the algebra may be differentially difficult while the reverse is true for the other group. It is ETS' practice to carry out DIF analysis on all tests they design in order to detect test items with differential performance for subgroups defined by gender and ethnicity.

The DIF program was developed at Educational Testing Service (Holland and Thayer, 1986) and was based on the Mantel-Haenszel odds-ratio (Mantel and Haenszel, 1959) and its associated Chi-Square. Basically, the Mantel-Haenszel (M-H) procedure forms odds ratios from two-way frequency tables. In a twenty item test, 21 two-way tables and their associated odds-ratios can be formed for each item. There are potentially 21 of these tables for each item since there will be one table associated with each total score from 0-20. The first dimension of each table is groups, e.g., Whites vs. Blacks, and the remaining dimension is passing vs. failing on a given item. Thus the question that the M-H procedure addresses itself to is whether or not members of the reference group, e.g., Whites, who have the same total score as members of the focal group, e.g., Blacks, have the same likelihood of passing the item in question. While the M-H statistic looks at passing rates for two groups while controlling for total score, no assumption need be made about the shape of the total score distribution for either group. The chi-square statistic associated with the M-H procedure tests whether the average odds-ratio for a test item, aggregated across all 21 score levels differs from unity, i.e., equal likelihood of passing.

The M-H procedure provides a statistical test of whether or not the average odds-ratio significantly departs from unity for each item. If the probability is .05 or less, then one could say that there is statistical evidence for DIF on the item in question. The problem with this interpretation is two-fold. First, one is making a large number of statistical tests, one for each item, so low probabilities will be found occasionally even if no DIF is present. Second, if there are two relatively large samples involved, statistical significance will be guaranteed.

Given these reservations, Educational Testing Service has developed an "effect size" estimate that is not sample size dependent. Associated with the effect sizes is a letter code that ranges from "A" to "C". It is ETS's experience that effect sizes of 1.5 and above have practical significance. Effect sizes of this magnitude, and which are statistically significant, are labelled with a "C". Items labelled "A" or "B" either do not show statistically significant differential functioning for the two groups being compared, or have differences that are too small to be important. Test development experts inspect items that are characterized by such large DIF properties, and in some cases are able to identify the reason, other than bias, for the differential item functioning.

If DIF statistics have been obtained on pretested items, <u>all</u> "C" items will normally be replaced in construction of an operational test, unless they are needed to meet test specifications. This is done

regardless of whether the group differences are related to the construct. Once a test has been administered, however, replacement of items is no longer an option; the only choice possible is whether to accept the questioned item or drop it from scoring. At this stage, it has been the policy of the Educational Testing Service to submit items having "C" level DIF statistics to a test development committee for review. If the committee can identify content that is likely to be unfamiliar to the subgroup in question and which is irrelevant to the skill being measured the item will typically be removed from the test score. However, if the identified source of difference is consistent with the construct being measured, or if no reason for the difference can be determined, the item is retained.

Table 3.2 presents a summary of the DIF results for the various subpopulations. The bottom of the table presents a summary of the number of "C" level DIF's accumulated across all content levels. Twenty-four items in total favored the reference groups while fifteen favored the focal groups. These two proportions do not differ significantly. This result, along with the fact that one might expect up to five percent occurrences by chance alone suggests that there is little potential DIF in the NELS:88 battery.

Speededness

Table 3.3 presents speededness indices for the gender, racial/ethnicity groups and totals. The speededness index presented here is the percentage of students in each group who attempt the last item. If over 80% attempt the last item the test is not assumed to be speeded, that is, differences in test performance are judged not to be due to time constraints. To a certain extent the proportion attempting the last item is at best an approximate estimate of speededness and likely to be biased in the direction of showing speededness when it is not present. One reason for this is that the items at the end of the test form tend to be the most difficult. As items near the end increase in difficulty, they may not be attempted by the less advanced students, and the speededness index would infer that the test is speeded rather than just having items towards the end that are too difficult for some test takers. Another reason for not answering one or more items at the end of the test might be lack of motivation to complete a test for which the student will be neither rewarded nor punished. Inspection of Table 3.3 suggests that there appears to be little problem with speededness. Not unexpectedly, speededness indices for the twelfth grade high math form fell below 80% for some subgroups. This form had five very difficult items at the very end. Another speededness index defines a test as not being speeded if "almost all" test takers complete 80% of the test. This definition is not affected by clusters of hard items at the end of the test. When this criterion was applied, the percentages completing at least 80% of the test exceeded 95% for virtually all subgroups and this finding was consistent for all grade levels. The vast majority of students who took the NELS:88 tests answered all of the questions. There is little indication that time constraints differentially affected scores for any gender or racial/ethnic subgroup.

Table 3.2
Counts of "C" Level DIF Items

Group Favored	Reading	Math	Science	History	Total
Base Year					
White (Reference Group)	0	0	0	1	1
Asian (Focal Group)	0	0	0	1	1
White (Reference Group)	0	0	0	0	0
Hispanic (Focal Group)	0	0	0	1	1
White (Reference Group)	0	1	1	0	2
Black (Focal Group)	0	0	0	1	1
Male (Reference Group)	0	1	0	1	2
Female (Focal Group)	0	0	0	0	0
First Follow-up					
White (Reference Group)	0	1	0	2	3
Asian (Focal Group)	0	0	0	1	1
White (Reference Group)	0	0	0	1	1
Hispanic (Focal Group)	0	0	0	1	1
White (Reference Group)	0	2	0	0	2
Black (Focal Group)	0	2	0	0	2
Male (Reference Group)	0	1	1	1	3
Female (Focal Group)	0	0	0	0	0
Second Follow-up					
White (Reference Group)	0	2	0	2	4
Asian (Focal Group)	1	1	0	3	5
White (Reference Group)	0	0	0	1	1
Hispanic (Focal Group)	0	0	0	1	1
White (Reference Group)	0	1	0	0	1
Black (Focal Group)	1	0	0	0	1
Male (Focal Group)	1	2	1	0	4
Female (Focal Group)	0	1	0	0	1

Table 3.2
Counts of "C" Level DIF Items (cont'd)

Summary	# Favoring Ref Group	# Favoring Focal Group	Total # C Items	Total Items in Pool	x 4 Contrasts	% of C-DIF Items
Base Year	5	3	8	116	464	1.7%
1st Follow-up	9	4	13	148	592	2.0%
2nd Follow-up	10	8	18	159	636	2.8%

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 3.3
Percentages of Selected Subgroups
Who Attempted the Last Item for Each Cognitive Test

	Total	Male	Female	Asian	Hispanic	Black	White
Base Year							
Reading	96%	95%	96%	96%	93%	90%	97%
Math	95%	95%	95%	96%	93%	90%	96%
Science	97%	97%	98%	97%	96%	94%	98%
History	98%	98%	98%	97%	97%	97%	99%
First Follow-up							
Reading Low	94%	95%	94%	92%	89%	90%	97%
Reading High	98%	98%	98%	97%	96%	93%	98%
Math Low	97%	97%	98%	99%	97%	96%	98%
Math Middle	94%	94%	94%	92%	90%	90%	96%
Math High	97%	97%	98%	98%	94%	96%	97%
Science	98%	98%	98%	96%	95%	96%	99%
History	98%	98%	97%	97%	95%	95%	98%

Table 3.3
Percentages of Selected Subgroups
Who Attempted the Last Item for Each Cognitive Test (cont'd)

	Total	Male	Female	Asian	Hispanic	Black	White
Second Follow-up							
Reading Low	93%	93%	93%	87%	87%	90%	95%
Reading High	91%	91%	91%	92%	83%	75%	93%
Math Low	98%	97%	98%	94%	96%	97%	99%
Math Middle	91%	92%	90%	91%	87%	87%	92%
Math High	81%	82%	79%	87%	69%	67%	82%
Science	97%	97%	97%	98%	95%	95%	98%
History	97%	97%	97%	95%	93%	95%	98%

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Motivation

The analysis above suggests that for those students who attempted the cognitive battery, motivation is not a problem. There is still a concern that those students who did not take the cognitive battery for whatever reason may not be missing at random particularly in the twelfth grade. Tables 3.4 and 3.5 present both the unweighted and weighted proportion of students who took cognitive tests in each content area, broken

Table 3.4
Percentage of Subgroups with Scorable Tests
Unweighted

Base Year	N	Reading	Math	Science	History
Total	16,489	96.3	96.3	96.2	95.9
Male	8,140	96.1	96.1	96.1	95.7
Female	8,349	96.5	96.4	96.3	96.1
Asian	976	96.9	96.5	96.4	96.0
Hispanic	2,010	94.7	94.4	94.4	94.2
Black	1,610	95.0	95.2	94.6	94.4
White	11,577	96.7	96.7	96.7	96.4
American Indian	162	98.8	98.8	98.8	98.8
Public	13,640	96.2	96.1	96.0	95.7
Catholic	1,308	97.0	97.2	97.2	97.0
NAIS Private	1,068	97.5	97.5	97.5	97.5
Other Private	473	96.2	96.4	96.2	95.1
Quartile					
SES Low	3,793	94.8	94.7	94.8	94.5
SES Second	3,908	96.1	96.0	96.1	95.7
SES Third	3,925	96.8	96.8	96.7	96.6
SES High	4,862	97.2	97.2	97.0	96.7

Table 3.4
Percentage of Subgroups with Scorable Tests
Unweighted (cont'd)

First Follow-up	N	Reading	Math	Science	History
Total	16,489	94.2	94.0	93.5	93.0
Male	8,140	93.9	93.7	93.2	92.7
Female	8,349	94.4	94.2	93.7	93.2
Asian	995	93.9	93.4	92.7	92.1
Hispanic	2,017	91.2	90.8	89.4	88.2
Black	1,628	92.0	91.5	90.8	90.0
White	11,662	95.0	94.9	94.6	94.3
American Indian	178	92.1	92.1	92.1	90.4
Public	13,594	95.9	95.7	95.2	94.6
Catholic	911	96.9	97.1	97.1	97.3
NAIS Private	966	93.5	93.3	92.7	92.0
Other Private	348	96.8	97.1	97.1	97.1
Quartile					
SES Low	3,671	90.9	90.4	89.3	88.7
SES Second	3,919	94.3	94.1	93.8	93.2
SES Third	3,980	95.2	95.1	94.8	94.3
SES High	4,918	95.6	95.6	95.3	94.9
In School	15,764	96.0	95.8	95.3	94.8
Dropout	631	53.9	52.9	52.1	52.3

Table 3.4
Percentage of Subgroups with Scorable Tests
Unweighted (cont'd)

Second Follow-up	N	Reading	Math	Science	History
Total	16,489	77.1	77.1	76.6	76.2
Male	8,140	77.2	77.2	76.7	76.2
Female	8,349	77.1	77.0	76.5	76.2
Asian	995	77.3	77.4	76.9	76.3
Hispanic	2,017	72.5	72.5	72.0	71.7
Black	1,628	73.1	73.1	72.1	71.6
White	11,662	78.6	78.6	78.2	77.8
American Indian	178	66.9	67.4	67.4	66.3
Public	12,585	81.5	81.5	80.9	80.5
Catholic	850	85.2	85.2	84.7	83.8
NAIS Private	930	78.8	78.9	78.8	78.8
Other Private	342	78.9	78.7	78.7	78.1
Quartile					
SES Low	3,663	71.9	71.9	71.3	70.8
SES Second	3,942	77.7	77.7	77.1	76.8
SES Third	4,024	78.4	78.3	77.8	77.4
SES High	4,859	79.6	79.6	79.2	78.9
In School	14,644	81.6	81.6	81.1	80.7
Dropout	1,116	41.8	41.3	41.0	40.9

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 3.5
Percentage of Subgroups with Scorable Tests
Weighted

Base Year	Wtd N	Reading	Math	Science	History
Total	2,970,835	96.2	96.2	95.9	95.6
Male	1,492,789	95.7	95.7	95.4	95.1
Female	1,478,047	96.8	96.6	96.3	96.2
Asian	102,531	96.5	95.9	95.2	95.2
Hispanic	306,232	95.0	94.6	94.5	94.3
Black	387,401	92.4	92.9	90.5	90.2
White	2,105,254	97.1	96.9	97.0	96.8
American Indian	36,415	99.3	99.3	99.3	99.3
Public	2,613,787	96.0	95.9	95.6	95.4
Catholic	224,755	97.5	97.7	97.7	97.5
NAIS Private	29,741	96.4	96.4	96.4	96.4
Other Private	102,552	98.5	98.6	98.4	98.3
Quartile					
SES Low	726,089	95.0	94.7	94.8	94.6
SES Second	733,914	96.1	96.2	96.2	95.8
SES Third	744,331	97.1	97.1	96.4	96.2
SES High	766,295	96.7	96.6	96.1	95.9

Table 3.5
Percentage of Subgroups with Scorable Tests
Weighted (cont'd)

First Follow-up	Wtd N	Reading	Math	Science	History
Total	2,970,835	91.8	91.5	91.0	90.7
Male	1,492,789	91.8	91.5	91.0	90.7
Female	1,478,047	91.8	91.6	91.0	90.8
Asian	105,878	91.9	91.4	90.8	90.4
Hispanic	307,485	87.9	87.6	86.3	85.2
Black	390,455	86.6	85.8	84.2	84.1
White	2,122,702	93.4	93.2	93.0	92.8
American Indian	42,530	90.6	91.4	91.5	90.1
Public	2,493,471	94.5	94.2	93.7	93.3
Catholic	168,244	95.3	95.0	95.0	95.5
NAIS Private	33,969	94.9	94.8	94.5	94.2
Other Private	75,608	91.6	91.7	91.7	91.7
Quartile					
SES Low	705,165	88.2	87.7	86.8	86.4
SES Second	734,788	90.9	90.6	90.1	89.7
SES Third	752,009	93.2	93.0	92.7	92.5
SES High	778,667	94.5	94.5	94.2	94.0
In School	2,767,772	94.5	94.3	93.9	93.5
Dropout	181,535	52.7	52.0	51.0	51.3

Table 3.5
Proportion of Subgroups with Scorable Tests
Weighted (cont'd)

Second Follow-up	Wtd N	Reading	Math	Science	History
Total	2,970,835	73.7	73.6	73.1	72.8
Male	1,492,789	74.4	74.4	73.8	73.5
Female	1,478,047	72.9	72.8	72.3	72.0
Asian	105,878	77.5	77.5	77.1	76.4
Hispanic	307,485	69.4	69.3	68.6	68.3
Black	390,455	67.6	67.6	66.9	66.8
White	2,122,702	75.4	75.3	74.8	74.5
American Indian	42,530	65.2	66.0	66.0	64.5
Public	2,253,702	79.8	79.7	79.1	78.8
Catholic	149,655	79.6	79.6	79.2	78.6
NAIS Private	32,107	78.8	78.8	78.6	78.8
Other Private	69,107	77.3	77.1	77.1	76.8
Quartile					
SES Low	702,256	67.7	67.7	66.9	66.4
SES Second	740,571	74.0	73.9	73.2	72.9
SES Third	756,102	74.7	74.6	74.2	74.2
SES High	771,700	77.9	77.8	77.5	77.0
In School	2,491,861	79.9	79.8	79.3	78.9
Dropout	301,788	42.4	42.1	41.7	41.7

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

out by subgroup within time point. Inspection of Tables 3.4 and 3.5 indicates that there is a dropoff in participation rates at the second follow-up. This decline in participation rates does not appear to be completely random. There is some indication that the lowest SES quartile was less likely to participate in the second follow-up cognitive testing. This apparent bias in response rates may lead to some bias in the estimates of the gain between the first and second follow-up. It is suggested here that researchers might estimate gain under differing assumptions about the causal mechanism underlying the missing scores to get a "handle" on the robustness of their population estimates. Checks on the robustness of one's estimates is desirable here since no attempt was made to develop test score sampling weights that are adjusted for non-response.

Table 4.1 in the next section compares the eligible NELS population of second follow-up grade 12 students with those who actually took the cognitive battery and also shows the comparable figures for the NAEP twelfth grade sample. (By definition, all NAEP participants took the NAEP tests. Students who were selected but for some reason not tested were deleted from the sample. However, NELS:88 sample members who were not tested may have participated in some other part of the survey, and remained in the sample.) These are weighted estimates. Table 4.1 indicates that about 78% of the eligible seniors took the cognitive battery, while 22% of the seniors did not take the cognitive battery. However, the subpopulation percentages of those who did participate reflect pretty much the same proportions as the second follow-up eligible population. There appears to be little evidence here suggesting that the missing cognitive scores for the in-school weighted population are non-representative of the eligible in-school population.

Chapter 4
Normative and Proficiency Level Scores

The cognitive test scores on the NELS:88 data files are of two broad types, normative scores and mastery scores. The normative scores are estimates of overall test performance and are available for all four cognitive areas at all three time points. Several transformations of the normative scores are included in the database: each of the scores is included in the original IRT-Estimated Number Right metric; each is transformed to a T-score metric, with standardization being done with respect to both the cross-sectional and longitudinal samples; finally, a quartile score ranks each test taker within the cross-sectional distribution of scores at each time point.

The second broad type of scores are mastery scores, or criterion referenced proficiency scores. These measure mastery of certain skill levels rather than being overall measures of performance. In the NELS:88 test battery, mastery levels have been defined only for the reading, math and science tests. Dichotomous and continuous measures of mastery are included in the database. The first is an indicator of whether the test taker passed or failed the cluster of test items that defined each proficiency level. The continuous measures represent the probability of a test taker passing each level, based on overall test performance.

Each of the scores in the database is discussed separately below.

IRT Estimated Number Right

The IRT-estimated number right for any individual at any one of the three time periods reflects an estimate of the number of items that a person would have answered correctly if he or she had taken all of the items that appeared in any form of the test. It is the probability of a correct answer on each item, summed over the total mathematics 81-item pool. The Bayesian Item Response Theory model allows one to put all the scores in, say Mathematics, on the same vertical scale so that the scores, regardless of the grade, can be interpreted in the same way. All the normal statistical operations that apply to any cognitive test score can be legitimately applied to the IRT-estimated number right. For example, a student's IRT-estimated number right in Mathematics in the tenth grade might be 41.3. That same student might have had an IRT-estimated number right of 35.3 in Math in the eighth grade and 44.5 in the twelfth grade. This particular student gained six points between the eighth and tenth grade (41.3 - 35.3 = 6) and 3.2 points between the tenth and twelfth grade (44.5 -41.3 = 3.2). The student's total gain over the four years was 9.2 points. The IRT-estimated number right in theory could range from a random guessing score to 81 correct in Mathematics. In fact, no one in the sample has either a random guessing score or a perfect score in Mathematics. The reader will notice that the IRT-estimated number right scores are not necessarily whole numbers, but typically include a decimal since they represent sums of probabilities. IRT scoring takes into consideration the pattern of correct answers and not just the simple number correct. In this sense IRT scoring tries to make use of all the information in the answer pattern. Everybody who has taken any test on any one or more of the three occasions will have at least one score in this metric. That is, an individual does not have to be a member of the longitudinal sample to have a score in this metric.

IRT Theta "T" Score

The IRT Theta "T" score has a mean of 50 and a standard deviation of 10 where the standardization (mean 50 and SD of 10) was carried out on the weighted panel sample, i.e., on people who were NELS:88 core sample participants in all three waves. As in the case of the IRT-estimated number right all individuals, regardless whether they were in the panel sample or not, will have a score in this metric for any time point(s) in which they did have a test score. The IRT-estimated number right is a non-linear transformation of the original theta scores. The rank ordering of individuals on this metric and the IRT-estimated number right metric is identical. As in the case of the IRT-estimated number right all the usual statistical operations that are typically used with gain scores are appropriate. Since the IRT-estimated number right is tied to the total item pool and thus the metric may seem more interpretable, one might prefer the IRT-estimated number right metric to the "T" score Theta metric. For example, an individual who has an estimated IRT-estimated number right of, say 40.3, can be said to be expected to get about half the items correct in the total pool. Because of the non-linear transformation between the Theta metric and the IRT-estimated number right metric the Theta metric tends to "stretch" out the scores at the extreme tails. This would have little impact on virtually all the typical statistical analysis done on gain scores and thus any analyses using the IRT-estimated number right or the Theta metric scores will be similar. The choice between the two is more a matter of preference of one metric or the other with respect to interpretability.

Cross-Sectional Scores

There are four additional cross-sectional scores available on the NELS:88 data files. These scores are called cross-sectional because they are all calibrated **within** each of the three separately-weighted sample waves. These cross-sectional scores are primarily used in statistical tables that describe score results within a particular grade, e.g., the twelfth grade, and use the cross-sectional weights associated with that wave of data.

Each of the four content areas in each of the three waves has a **t-score transformation of the IRT Estimated Number Right score**. Unlike the Theta t-score, which is standardized with respect to all three waves of data combined, this transformation is based on the test scores for each year considered separately. All scores for core (weighted) sample members, including freshened samples in the two follow-up years, are used in obtaining the parameters for the transformation to a mean of 50 and SD of 10. That is, the IRT Estimated Number Right T Score will have this weighted mean and standard deviation when aggregated **over all core participants in a single year with the cross-sectional weight used in computing the statistics**. Test takers who are not in the weighted core sample also have this score, which is computed using the same parameters as the core sample, but will not necessarily result in the same mean and standard deviation.

All four content areas in each of the three grades have **Achievement Quartile** scores, which are based on a weighted frequency distribution of core sample students within each year. The IRT Number Right Score, IRT t-score, and Theta t-score all preserve the same rank-ordering of students within year. Any of these can be used to determine the score cut points that divide the weighted frequency distribution into four equal groups. A quartile score of "1" corresponds to the lowest group, and "4" is the highest. Quartile scores are also assigned to test takers who are not in the core sample by using the same cut points as for the core students. The appropriate interpretation of a quartile score of "2" for an augmented-sample student in the second follow-up, for example, would be: "This student has a score that would put him or

her in the second quartile of twelfth graders nationwide in that year." Again, quartile scores for additional samples will not necessarily divide the other samples into four equal groups, since the distribution of scores may not match that of the nationally representative weighted core sample.

Each test taker who has a reading score and/or a math score also has a **Reading+Math Composite T-Score**. This is the equally-weighted average of the standardized (t-metric) reading and math scores, with one or the other used alone if one is missing. The reading and math IRT Estimated Number Right scores have different means and standard deviations, so the transformed scores are used for building the composite in order to give equal weight to both subject areas. The composite is then re-standardized, again within the core sample for each wave and using the cross-sectional weights, to produce a score that has a mean of 50 and SD of 10 when aggregated for this group. The weighted frequency distribution of the composite is divided into four equal groups for the **Reading+Math Composite Quartile** score. As described above, the parameters for standardizing the composite and the cut points for dividing it into quartiles are also applied to the non-core samples to produce scores that allow these samples to be compared to national population estimates.

Criterion-Referenced Proficiency Scores

In addition to the normative interpretations in the NELS cognitive tests, the reading, mathematics, and science tests also provide criterion referenced interpretations. The criterion- referenced interpretations are based on students demonstrating proficiencies on clusters of items that mark ascending points on the test score scale. For example, there are three separate clusters of items in reading that mark the low, middle, and high end of the reading scale. The items that make up these clusters exemplify the skills required to successfully answer the typical item located at these points along the scale.

General Description of the Proficiency Levels

The three levels of proficiency in the reading test, five in the mathematics test, and three in the science test, are as follows:

Reading

Reading Level 1: Simple reading comprehension including reproduction of detail and/or the author's main thought.

Reading Level 2: Ability to make relatively simple inferences beyond the author's main thought and/or understand and evaluate relatively abstract concepts.

Reading Level 3: Ability to make complex inferences or evaluative judgments that require piecing together multiple sources of information from the passage.

Mathematics

Math Level 1: Simple arithmetical operations on whole numbers: essentially single step operations which rely on rote memory.

Math Level 2: Simple operations with decimals, fractions, powers and roots.

Math Level 3: Simple problem solving, requiring the understanding of low level mathematical concepts.

Math Level 4: Understanding of intermediate level mathematical concepts and/or having the ability to formulate multi-step solutions to word problems.

Math Level 5: Proficiency in solving complex multi-step word problems and/or the ability to demonstrate knowledge of mathematics material found in advanced mathematics courses.

Science

Science Level 1: Understanding of everyday science concepts; "common knowledge" that can be acquired in everyday life.

Science Level 2: Understanding of fundamental science concepts upon which more complex science knowledge can be built.

Science Level 3 Understanding of relatively complex scientific concepts; typically requiring an additional problem solving step.

There are two kinds of criterion referenced proficiency scores. The first kind is a **dichotomous** score of "0" or "1" where a "1" indicates mastery of the material at this objective level and a "0" implies non-mastery. The second kind is a continuous score indicating the **probability** that a student has mastered the type of items that describe a particular criterion referenced level. The proficiency levels are hierarchically ordered in the sense that mastery of the highest level among three levels implies that one would have also mastered the lower two levels. A student who has mastered all three hierarchical levels would have a dichotomous score pattern for the three levels of [1 1 1]. Similarly a student who only mastered the first two levels would have a dichotomous score pattern of [1 1 0]. A "reversal" pattern such as [0 11], that is, a failed easy level followed by one or more passed more difficult levels, is inconsistent with the hierarchical model. Students who omitted items that were critical to determining proficiency level, or who have reversals in proficiency score patterns will have a "blank" instead of a "0" or "1". Students who took enough of the items marking the proficiency levels **and** who had no reversals will have "0" or "1" scores for each of the proficiency levels that were available for that grade and content area. The vast majority of students did fit the hierarchical proficiency model, i.e., had no reversals. Dichotomous proficiency scores are present for reading, mathematics, and science. The twelfth grade had typically more dichotomously scored proficiency levels than the lower grades since it always incorporated all the lower levels plus any new more difficult level(s). Also the most difficult mathematics form did not include the easiest proficiency level and the easiest form did not include the most difficult proficiency level. There were four items that served as markers for each proficiency level. A student was defined to be proficient at a given proficiency level if he or she got any 3 of 4 items correct that "mark" that level. Items were selected for a proficiency level if they shared similar cognitive processing requirements and this cognitive demand similarity was reflected in similar item difficulties.

Analyses using the dichotomous proficiency scores include descriptive statistics that show the percentages of various subpopulations who have demonstrated proficiencies at each of the hierarchical

levels. They can also be used to examine patterns of change with respect to proficiency levels. An example of this type of analysis using dichotomous proficiency scores can be found in Rock, Owings & Lee (1994).

The second kind of proficiency score is the **probability** of being proficient at each of the levels. This is a continuous analog to the dichotomous proficiency scores. The advantage of the **probability** of being proficient at each of the levels over the dichotomous proficiencies is that: (1) They are continuous scores and thus all the more powerful statistical methods can be applied, and (2) probabilities of being proficient at each of the levels, say in grade 10 are available for any individual who had a test score in grade 10. This second advantage is true since the IRT model enables us to estimate how a person would do on even those items that he or she was not given, e.g., if they were on a different form or not given in that grade. By contrast, the item-based dichotomous scores depend heavily on students answering the actual items in the cluster.

The proficiency probabilities are particularly appropriate for relating specific processes to changes that occur at different points along the score scale. Since the proficiency levels are hierarchical they mark different ascending points along the score scale. For example, one might wish to evaluate the impact of taking advanced math courses on changes in mathematics from grade 10 to grade 12. One approach to doing this would be to subtract every student's tenth grade IRT-estimated number right from the their twelfth grade IRT-estimated number right and correlate this difference with the number of advanced mathematics courses taken between the tenth and twelfth grade. The resulting correlation will be relatively small because individuals taking no advanced mathematics courses are also gaining but probably at the low end of the test score scale. Individuals who are taking advanced mathematics courses are also gaining but at the higher end of the test score scale. To be more concrete, let us say that the individuals who took none of the advanced math courses gained on average 3 points, all at the low end of the test score scale. Conversely the individuals who took the advanced math courses gained 4.5 points but virtually all these individuals made their gains at the upper end of the test score scale. When the researcher correlates courses with gains, the fact that on average the advanced math takers gained only slightly more than those taking no advanced mathematics courses will lead to a very small correlation between gain and process (advanced math course taking). This low correlation has nothing to do with reliability of gain scores, but it has much to do with **where** on the test score scale the gains are taking place. Gains in the upper end of the test score distribution reflect increases in knowledge in advanced mathematical concepts and processes while gains at the lower end reflect gains in basic arithmetical concepts. In order to relate specific processes to gains successfully one has to match the process of interest to where the gain is taking place.

The proficiency probabilities do this since they mark ascending places on the test score distribution. If I wish to relate the number of advanced math courses taken to changes, I should be looking at changes at the upper end of the test score distribution. How does one use the proficiency probabilities to do this? There are five proficiency levels in mathematics with level 4 and level 5 marking the two highest points along the test score scale. One would expect the taking of advanced math courses to have its greatest effects on changes in probabilities of being proficient at these highest two levels. Thus one would simply subtract each individuals tenth grade probability of being proficient at say level 4 from the corresponding probability of being proficient at level 4 in twelfth grade. Now every individual has a continuous measure of change in mastery of advanced skills rather than along the whole score scale. One then correlates this change in level 4 probabilities with the number of advanced mathematics courses taken and we will observe a substantial increase in the relationship between change and process (number of advanced mathematics courses taken). One might wish to do the same thing with the level 5 probabilities as well. The main point here is that certain school processes, in particular, course taking

patterns, target gains at different points along the test score distribution. One has to match the type of school process one is evaluating with the location on the test score scale where the gains are likely to be taking place and then select the proper proficiency levels for appropriately evaluating that impact. (For an example of the use of probability of proficiency scores to measure mathematics achievement gain in relation to program placement and course taking, see Chapter 4 of Scott, Rock, Pollack & Ingels, 1995).

NAEP Equated Score

The goals set out for the NELS:88 test battery in the base year included generation of mathematics cross-walks with two other studies. The NELS:88 tests were to share sufficient common items with the HS&B battery to support cross-sectional equating with the 1980 HS&B sophomore cohort in mathematics (for an example of such HS&B/NELS:88 equating, see Rasinski, Ingels, Rock & Pollack, 1993). The NELS:88 tests were also to provide sufficient item overlap with the National Assessment of Educational Progress (NAEP) mathematics test at twelfth grade to cross-walk to the NAEP mathematics scale.

Hence a score on the NAEP scale in mathematics has been placed on the NELS:88 1992 data file for every student who had a twelfth grade NELS mathematics score. This is an equated score based on an equipercentile equating procedure. The validity of the equating procedure relies on the fact that both the NAEP and NELS samples are probability samples from the same parent population. In addition, the equating assumes that the test provided a reasonable match in content. Table 4.1 contains the subpopulation makeup of the two samples.

Table 4.1
A Comparison of the NAEP and NELS 12th Grade Samples

Estimated proportion of selected subpopulation based on weighted percentages			
	NAEP	**NELS Population**	**NELS Test Takers**
Total Population Estimate	2,522,170	2,537,024	1,979,737
Male	48.8%	50.4%	50.9%
Female	51.2%	49.6%	49.1%
White	71.1%	72.3%	73.3%
Black	14.7%	11.9%	11.4%
Hispanic	9.5%	10.0%	9.8%
Public	87.1%	89.9%	90.1%
Private	4.5%	4.3%	3.9%
Catholic	8.4%	5.8%	5.9%

Source: National Education Longitudinal Study of 1988: Second Follow-Up and National Assessment of Educational Progress 1992 Twelfth Grade Sample, U.S. Department of Education, National Center for Education Statistics.

Empirical checks on the validity of the equating procedure included comparing subgroup differences on the equated score with those found on the original NAEP scale. Virtually all checks were within one standard error. A researcher who wishes to look at the relationship between the background and process variables from the NELS data base using the NAEP mathematics scale score can now do so.

Chapter 5
Psychometric Properties of the NELS:88 Scores

In the final analysis the reliability and validity of the NELS:88 cognitive scores depend on the: 1) appropriateness of the test content specifications, 2) psychometric quality of the test items themselves, 3) appropriateness of the difficulty of the tests for the students being measured, 4) lack of speededness, 5) success of the IRT procedures used for linking across grades and forms, and 6) scoring procedures. Previous sections discussed content specifications, psychometric qualities of the items, appropriateness of item difficulties, speededness and linking procedures used. This chapter provides both traditional indices of reliability as well as IRT centered estimates. In addition evidence for the construct and predictive validity of the NELS:88 scores are presented.

Reliability of the IRT Scores

An approximate index of the reliability of the IRT theta estimates is presented in Table 5.1 by grade and content area. While the plot of the information function is the most comprehensive measure of the reliability of the IRT scores, it is sometimes helpful to present an estimate of the more familiar single index type. These indices are computed as 1 minus the ratio of the average measurement error variance to the total variance (see for example, Samejima, 1994).

Table 5.1
Reliability of Theta

	Base Year	First Follow-up	Second Follow-up
Reading	.80	.86	.85
Math	.89	.93	.94
Science	.73	.81	.82
History/Citizenship/Geography	.84	.85	.85

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

$$\gamma_{\alpha} = 1 - \frac{(\sum_{i=1}^{n} \sigma_{ei}^2)/N}{\sigma^2(\hat{\theta})}$$

where :

σ_{ei}^2 = posterior variance for the *i*th subtest

$\sigma^2(\hat{\theta})$ = variance of the thetas

Inspection of Table 5.1 indicates that the introduction of the adaptive forms in grade 10 and 12 reading and math, lead to substantial increases in reliability. It should be noted that the base year psychometric report (Rock & Pollack, 1991) reported coefficient alpha reliabilities based on the observed scores. Because of the adaptive nature of the reading and mathematics tests at first and second follow-up the same reliability estimation procedure was no longer appropriate. This report, in order to be consistent across all subject areas and time points, used the IRT reliability estimation procedure for all measures whether they were adaptive or not. The information functions are presented in Appendix G. The test information function shows the relationship between the amount of information available in the items for estimating the ability scores at each point in the ability distribution. More specifically, the test information function estimates the reciprocal of the squared standard error of measurement at each ability level. The greater the amount of information at a given ability level, the more closely the estimates of ability cluster around the true ability level (Baker, 1992). That is, the greater the height of the test information function the more precise the estimates. The fact that the height of the curve is much reduced as one moves towards the tails indicates that the maximum information function occurs in the middle of the range, where the item difficulty approximates the abilities of the majority of the test takers. This latter property is precisely why the NELS:88 battery developed adaptive test forms in mathematics and reading.

Construct Validity of the NELS:88 Content Areas

Table 5.2 presents the intercorrelations of the content areas by year of administration. There is some tendency for the intercorrelations among content areas to increase with grade in school. That is the average intercorrelations among content areas are .72, .75, and .76 for the eighth, tenth, and twelfth grade respectively. Correlations between adjacent administrations within the same content areas tend to be higher then those found between content areas within the same administration. The finding is consistent with the notion that the content areas should show some discriminant validity. Additional information on the discriminant validity for the content areas can be found in Rock & Pollack (1991). Also correlations between eighth and tenth grade scores tend to be lower than those found between tenth and twelfth grade scores within all the content areas. This is consistent with the fact that proportionately greater changes in achievement measured by these tests occurred between the eighth and tenth grade than occurred between the tenth and twelfth grade.

While the internal correlational analyses among the scale scores show some discriminant and convergent validity for the content areas, they tell us little about how well the application of Bayesian IRT approaches "worked" compared to the more traditional baseline technique based on the LOGIST conditional maximum likelihood estimation. The following discussion presents some results comparing two variations of the Bayesian approach with each other and with LOGIST. The results are presented for the mathematics content area since it was the most complex to scale because of its seven forms. Validity for the three approaches to IRT scaling as well as for the content areas themselves is defined here in terms of the pattern of correlations between their IRT scores and relevant outside process and demographic variables. In the end longitudinal studies that emphasize policy decisions must concern themselves with describing the extent of the relationship between student performance and school and home-based learning experiences.

Table 5.2
Intercorrelations of Content Areas
Within and Across Administrations

	READ BY	MATH BY	SCI BY	HIST BY	READ F1	MATH F1	SCI F1	HIST F1	READ F2	MATH F2	SCI F2	HIST F2
READ BY	1.00											
MATH BY	0.71	1.00										
SCI BY	0.71	0.73	1.00									
HIST BY	0.73	0.69	0.73	1.00								
READ F1	0.80	0.69	0.68	0.71	1.00							
MATH F1	0.69	0.88	0.70	0.67	0.76	1.00						
SCI F1	0.66	0.72	0.74	0.68	0.74	0.79	1.00					
HIST F1	0.67	0.65	0.68	0.70	0.75	0.72	0.77	1.00				
READ F2	0.74	0.65	0.64	0.66	0.82	0.71	0.69	0.70	1.00			
MATH F2	0.66	0.83	0.68	0.65	0.73	0.92	0.77	0.70	0.74	1.00		
SCI F2	0.63	0.70	0.71	0.65	0.69	0.75	0.80	0.70	0.73	0.79	1.00	
HIS F2	0.66	0.64	0.66	0.71	0.71	0.69	0.72	0.78	0.75	0.73	0.77	1.00

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

One of the concerns outlined above in the preceding scaling chapter was the potential for LOGIST estimates to have ceiling effects for high scoring tenth grade students. Such students would not have any "room" to gain between the tenth and twelfth grades. We would expect that such limiting effects if they exist would show up when groups of advanced students were compared with groups of students who are less advanced. For example, one might get an underestimate of differences in gains between the students who take advanced mathematics courses versus those who do not. Part of this underestimate may be attributable to the fact that LOGIST procedures have no systematic way to deal with ceiling and near ceiling effects for high scoring students on the base year and first follow-up tests.

Tables 5.3, 5.4 and 5.5 present correlations of gains and selected background and process variables. Gains are shown in the Theta and "true" score metric for the 8-10, 10 - 12, and the 8 - 12 (total gain) for LOGIST estimates and for two kinds of Bayesian approaches (ST1 and ST4). In addition, grade 8 to 12 gains in proficiency probabilities at each of the five mathematics proficiency levels are also correlated with background and process variables. As indicated in Chapter 4 the proficiency probabilities are simply the probability that a given individual has "mastered" the skills defined by the items marking each of the proficiency levels. Like any score these probabilities can be monitored for gains taking place at any one of five proficiency levels. The Theta metric and the "true" score metric are also discussed in chapter 4. The two kinds of Bayesian procedures differ in whether they use a normal prior (ST1) or a distribution free prior (ST4).

Table 5.3
Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math
CORRELATIONS OF GAINS AND GRADE 12 STATUS WITH BACKGROUND VARIABLES
3 METHODS: "LOG"=LOGIST; "ST1" = NALS 1-STEP METHOD; "ST4" = NAEP 4-STEP METHOD

	Self-Reported Courses		Gender, Ethnicity, SES				School Type		Curriculum
	Any Math Last 2 YR	Taking Math Now	Male=1 Female=0	Hisp=1 White=0	Black=1 White=0	SES Quartile	Cath=1 Public=0	NAIS=1 Public=0	ACAD CUR=1 GEN/VOC=0
GAIN IN THETA METRIC									
GAIN 8-10 LOG	0.07	0.06	0.01	0.00	-0.01	0.05	0.03	0.01	0.06
GAIN 8-10 ST1	0.11	0.11	0.02	-0.04	-0.07	0.14	0.05	0.04	0.15
GAIN 8-10 ST4	0.08	0.06	0.02	-0.01	-0.04	0.07	0.03	-0.02	0.07
GAIN 10-12 LOG	0.07	0.15	0.05	0.01	-0.01	0.04	0.03	0.01	0.06
GAIN 10-12 ST1	0.14	0.23	0.08	0.00	-0.02	0.10	0.05	0.06	0.14
GAIN 10-12 ST4	0.10	0.18	0.07	0.03	0.03	0.02	0.03	0.01	0.06
TOTAL GAIN LOG	0.12	0.18	0.06	0.01	-0.02	0.08	0.05	0.02	0.11
TOTAL GAIN ST1	0.19	0.26	0.07	-0.04	-0.07	0.19	0.07	0.08	0.22
TOTAL GAIN ST4	0.14	0.18	0.06	0.02	-0.02	0.07	0.05	-0.01	0.10
GAIN IN TRUE SCORE METRIC									
GAIN 8-10 LOG	0.10	0.09	0.02	-0.03	-0.06	0.11	0.05	0.01	0.12
GAIN 8-10 ST1	0.11	0.11	0.02	-0.06	-0.09	0.15	0.05	0.03	0.16
GAIN 8-10 ST4	0.11	0.10	0.02	-0.06	-0.09	0.14	0.06	0.01	0.15
GAIN 10-12 LOG	0.12	0.18	0.06	0.00	-0.02	0.06	0.04	0.00	0.09
GAIN 10-12 ST1	0.14	0.21	0.07	0.01	-0.02	0.08	0.05	0.02	0.11
GAIN 10-12 ST4	0.14	0.22	0.07	0.01	-0.02	0.08	0.05	0.02	0.11
TOTAL GAIN LOG	0.18	0.22	0.06	-0.03	-0.07	0.14	0.08	0.01	0.17
TOTAL GAIN ST1	0.19	0.24	0.06	-0.04	-0.09	0.18	0.08	0.04	0.21
TOTAL GAIN ST4	0.19	0.23	0.06	-0.04	-0.09	0.17	0.08	0.02	0.20

Table 5.3 (cont'd)
Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math
CORRELATIONS OF GAINS AND GRADE 12 STATUS WITH BACKGROUND VARIABLES
3 METHODS: "LOG"=LOGIST; "ST1" = NALS 1-STEP METHOD; "ST4" = NAEP 4-STEP METHOD

	Self-Reported Courses		Gender, Ethnicity, SES				School Type		Curriculum
	Any Math Last 2 YR	Taking Math Now	Male=1 Female=0	Hisp=1 White=0	Black=1 White=0	SES Quartile	Cath=1 Public=0	NAIS=1 Public=0	ACAD CUR=1 GEN/VOC=0
GAIN IN PROFICIENCY PROBABILITY (8-12)									
GPL1 LOG	-0.05	-0.05	0.01	0.13	0.16	-0.19	-0.04	-0.08	-0.18
GPL1 ST1	-0.07	-0.09	0.00	0.16	0.20	-0.25	-0.05	-0.12	-0.24
GPL1 ST4	-0.07	-0.08	0.01	0.15	0.18	-0.23	-0.05	-0.10	-0.22
GPL2 LOG	0.05	-0.01	0.00	0.08	0.05	-0.12	-0.01	-0.14	-0.11
GPL2 ST1	0.05	-0.02	-0.01	0.07	0.04	-0.11	0.00	-0.14	-0.10
GPL2 ST4	0.01	-0.06	0.00	0.11	0.08	-0.17	-0.02	-0.15	-0.16
GPL3 LOG	0.14	0.13	0.03	-0.05	-0.10	0.12	0.07	-0.06	0.15
GPL3 ST1	0.13	0.10	0.02	-0.04	-0.09	0.09	0.06	-0.07	0.12
GPL3 ST4	0.13	0.10	0.03	-0.04	-0.09	0.09	0.06	-0.08	0.11
GPL4 LOG	0.17	0.30	0.06	-0.14	-0.17	0.34	0.10	0.20	0.35
GPL4 ST1	0.18	0.27	0.05	-0.15	-0.18	0.34	0.10	0.17	0.35
GPL4 ST4	0.17	0.31	0.06	-0.15	-0.17	0.36	0.10	0.23	0.37
GPL5 LOG	0.02	0.06	0.03	-0.03	-0.03	0.07	0.02	0.08	0.06
GPL5 ST1	0.08	0.18	0.07	-0.08	-0.09	0.23	0.03	0.24	0.20
GPL5 ST4	0.06	0.14	0.07	-0.06	-0.07	0.18	0.02	0.19	0.15

Table 5.4
Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math
CORRELATIONS OF GAINS AND GRADE 12 STATUS WITH MATH COURSES TAKEN
3 METHODS: "LOG"=LOGIST; "ST1" = NALS 1-STEP METHOD; "ST4" = NAEP 4-STEP METHOD

	# Units	Ave. Grade	Algebra1	Algebra2	Geometry	Trig	Pre-Calc	Calculus	Other Math
GAIN IN THETA METRIC									
GAIN 8-10 LOG	0.12	0.10	0.04	0.08	0.10	0.03	0.06	0.05	-0.08
GAIN 8-10 ST1	0.25	0.20	0.04	0.19	0.23	0.11	0.14	0.12	-0.20
GAIN 8-10 ST4	0.18	0.08	0.13	0.13	0.17	0.04	0.04	0.01	-0.16
GAIN 10-12 LOG	0.09	0.05	0.00	0.04	0.05	0.05	0.06	0.05	0.00
GAIN 10-12 ST1	0.21	0.16	-0.01	0.11	0.14	0.10	0.16	0.16	-0.06
GAIN 10-12 ST4	0.10	0.05	0.02	0.03	0.04	0.04	0.07	0.07	0.01
TOTAL GAIN LOG	0.20	0.14	0.04	0.12	0.14	0.08	0.11	0.09	-0.08
TOTAL GAIN ST1	0.35	0.28	0.03	0.23	0.29	0.16	0.23	0.21	-0.20
TOTAL GAIN ST4	0.22	0.11	0.12	0.14	0.18	0.07	0.09	0.05	-0.13
GAIN IN TRUE SCORE METRIC									
GAIN 8-10 LOG	0.22	0.15	0.08	0.18	0.22	0.09	0.09	0.05	-0.19
GAIN 8-10 ST1	0.27	0.20	0.07	0.22	0.27	0.12	0.13	0.08	-0.24
GAIN 8-10 ST4	0.26	0.18	0.09	0.22	0.27	0.11	0.11	0.06	-0.24
GAIN 10-12 LOG	0.15	0.06	0.05	0.09	0.11	0.07	0.07	0.03	-0.04
GAIN 10-12 ST1	0.19	0.10	0.05	0.10	0.14	0.08	0.10	0.08	-0.07
GAIN 10-12 ST4	0.20	0.10	0.05	0.11	0.15	0.08	0.10	0.08	-0.08
TOTAL GAIN LOG	0.31	0.18	0.11	0.23	0.28	0.13	0.13	0.07	-0.20
TOTAL GAIN ST1	0.36	0.23	0.09	0.26	0.32	0.16	0.17	0.12	-0.24
TOTAL GAIN ST4	0.35	0.22	0.11	0.26	0.32	0.15	0.16	0.10	-0.25

Table 5.4 (cont'd)
Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math
CORRELATIONS OF GAINS AND GRADE 12 STATUS WITH MATH COURSES TAKEN
3 METHODS: "LOG"=LOGIST; "ST1" = NALS 1-STEP METHOD; "ST4" = NAEP 4-STEP METHOD

	# Units	Ave. Grade	Algebra1	Algebra2	Geometry	Trig	Pre-Calc	Calculus	Other Math
GAIN IN PROFICIENCY PROBABILITY (8-12)									
GPL1 LOG	-0.20	-0.21	0.06	-0.19	-0.22	-0.14	-0.14	-0.13	0.21
GPL1 ST1	-0.26	-0.28	0.11	-0.25	-0.29	-0.19	-0.20	-0.18	0.25
GPL1 ST4	-0.25	-0.26	0.08	-0.23	-0.28	-0.17	-0.18	-0.16	0.25
GPL2 LOG	-0.02	-0.20	0.30	-0.03	0.00	-0.12	-0.20	-0.21	-0.04
GPL2 ST1	-0.01	-0.20	0.30	-0.02	0.03	-0.11	-0.20	-0.22	-0.07
GPL2 ST4	-0.08	-0.25	0.30	-0.10	-0.07	-0.16	-0.22	-0.23	0.01
GPL3 LOG	0.25	0.09	0.14	0.26	0.31	0.13	0.02	-0.10	-0.24
GPL3 ST1	0.22	0.05	0.17	0.23	0.29	0.10	-0.02	-0.13	-0.23
GPL3 ST4	0.22	0.04	0.19	0.23	0.29	0.09	-0.03	-0.14	-0.23
GPL4 LOG	0.44	0.48	-0.20	0.32	0.36	0.30	0.42	0.37	-0.25
GPL4 ST1	0.44	0.46	-0.17	0.36	0.39	0.31	0.39	0.29	-0.27
GPL4 ST4	0.46	0.52	-0.23	0.33	0.37	0.31	0.46	0.43	-0.26
GPL5 LOG	0.08	0.14	-0.07	0.02	0.03	0.02	0.10	0.19	-0.02
GPL5 ST1	0.25	0.38	-0.23	0.09	0.14	0.13	0.33	0.53	-0.09
GPL5 ST4	0.19	0.31	-0.18	0.07	0.11	0.09	0.26	0.42	-0.07

Table 5.5
Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math
CORRELATIONS OF GAIN WITH INITIAL (GRADE 8) STATUS
3 METHODS: "LOG"=LOGIST; "ST1" = NALS 1-STEP METHOD; "ST4" = NAEP 4-STEP METHOD

	THETA METRIC			TRUE SCORE METRIC		
	TH8 LOG	TH8 ST1	TH8 ST1	NR8 LOG	NR8 ST1	NR8 ST4
GAIN IN THETA METRIC						
GAIN 8-10 LOG	-0.2977	-0.1737	-0.1800	-0.1794	-0.1458	-0.1418
GAIN 8-10 ST1	-0.0465	-0.0106	-0.0080	-0.0171	-0.0043	-0.0076
GAIN 8-10 ST4	-0.1816	-0.1630	-0.1595	-0.1796	-0.1674	-0.1763
GAIN 10-12 LOG	-0.0074	0.0013	-0.0004	0.0043	0.0061	0.0070
GAIN 10-12 ST1	0.0520	0.0563	0.0512	0.0669	0.0634	0.0696
GAIN 10-12 ST4	-0.1164	-0.1115	-0.1194	-0.0935	-0.0960	-0.0855
TOTAL GAIN LOG	-0.2957	-0.1680	-0.1754	-0.1710	-0.1368	-0.1322
TOTAL GAIN ST1	0.0000	0.0321	0.0305	0.0345	0.0422	0.0441
TOTAL GAIN ST4	-0.2403	-0.2207	-0.2234	-0.2221	-0.2134	-0.2135
GAIN IN TRUE SCORE METRIC						
GAIN 8-10 LOG	-0.1147	-0.0742	-0.0667	-0.0998	-0.0795	-0.0901
GAIN 8-10 ST1	0.0116	0.0274	0.0379	0.0040	0.0158	0.0036
GAIN 8-10 ST4	0.0071	0.0188	0.0323	-0.0170	-0.0020	-0.0217
GAIN 10-12 LOG	0.0182	0.0166	0.0189	0.0126	0.0135	0.0106
GAIN 10-12 ST1	0.0046	-0.0018	-0.0020	-0.0012	-0.0039	-0.0030
GAIN 10-12 ST4	0.0048	-0.0004	-0.0007	0.0005	-0.0015	-0.0005
TOTAL GAIN LOG	-0.0872	-0.0526	-0.0441	-0.0784	-0.0597	-0.0714
TOTAL GAIN ST1	0.0128	0.0212	0.0297	0.0024	0.0103	0.0008
TOTAL GAIN ST4	0.0091	0.0153	0.0262	-0.0137	-0.0026	-0.0183

Table 5.5 (cont'd)
Evaluation of Alternative Scoring Procedures for Grade 8-10-12 Math
CORRELATIONS OF GAIN WITH INITIAL (GRADE 8) STATUS
3 METHODS: "LOG"=LOGIST; "ST1" = NALS 1-STEP METHOD; "ST4" = NAEP 4-STEP METHOD

	THETA METRIC			TRUE SCORE METRIC		
	TH8 LOG	TH8 ST1	TH8 ST1	NR8 LOG	NR8 ST1	NR8 ST4
GAIN IN PROFICIENCY PROBABILITY (8-12)*						
GAIN: LEVEL 1 LOG	-0.5979	-0.5595	-0.5856	-0.5067	-0.5025	-0.4700
GAIN: LEVEL 1 ST1	-0.6479	-0.6560	-0.6831	-0.6061	-0.6123	-0.5837
GAIN: LEVEL 1 ST4	-0.6611	-0.6158	-0.6447	-0.5545	-0.5515	-0.5159
GAIN: LEVEL 2 LOG	-0.4948	-0.5704	-0.5768	-0.5715	-0.5877	-0.5868
GAIN: LEVEL 2 ST1	-0.4461	-0.5355	-0.5330	-0.5520	-0.5703	-0.5772
GAIN: LEVEL 2 ST4	-0.5419	-0.6181	-0.6294	-0.6128	-0.6299	-0.6264
GAIN: LEVEL 3 LOG	-0.0601	-0.0992	-0.0652	-0.1509	-0.1475	-0.1717
GAIN: LEVEL 3 ST1	-0.0724	-0.1173	-0.0817	-0.1710	-0.1694	-0.1939
GAIN: LEVEL 3 ST1	-0.1353	-0.1921	-0.1588	-0.2458	-0.2472	-0.2721
GAIN: LEVEL 4 LOG	0.3666	0.4370	0.4470	0.4154	0.4448	0.4277
GAIN: LEVEL 4 ST1	0.3263	0.3846	0.4016	0.3567	0.3848	0.3652
GAIN: LEVEL 4 ST4	0.4002	0.4752	0.4843	0.4535	0.4835	0.4662
GAIN: LEVEL 5 LOG	0.4470	0.5406	0.5240	0.5449	0.5659	0.5669
GAIN: LEVEL 5 ST1	0.5232	0.6299	0.6065	0.6256	0.6484	0.6473
GAIN: LEVEL 5 ST4	0.5044	0.5899	0.5611	0.5967	0.6054	0.6139
GRADE 12 THETA AND TRUE SCORE						
GR12 THETA LOG	0.7593	0.8038	0.8017	0.7990	0.8020	0.7976
GR12 THETA ST1	0.7902	0.8440	0.8412	0.8390	0.8445	0.8397
GR12 THETA ST4	0.7855	0.8339	0.8346	0.8221	0.8284	0.8200
GR12 TRUE SCORE LOG	0.7700	0.8241	0.8238	0.8157	0.8229	0.8162
GR12 TRUE SCORE ST1	0.7850	0.8414	0.8407	0.8327	0.8406	0.8337
GR12 TRUE SCORE ST4	0.7864	0.8411	0.8423	0.8347	0.8424	0.8356

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Inspection of Tables 5.3, 5.4 and 5.5 indicates that in the Theta metric the normal prior Bayesian procedure (ST1) shows stronger relationships between gains and virtually all the process/demographic variables than do the other two procedures. The differences in favor of ST1 are particularly strong where contrasts are being made between groups quite different in their mathematics preparation, e.g., the relationship between being in the academic curriculum or "taking math now" and total gain.

When the correlations are based on the "true" score metric the ST1 Bayesian approach still does as well or better than the other two approaches. The "true" score metric is a non-linear transformation of the Theta scores and unlike the Thetas does not quite stretch out the tails of the score distribution as much as the Thetas. The stretching out at the tails has little impact on most analyses except if one is contrasting groups whose scores put them in or near the tail of the distribution.

The proficiency probabilities recorded in Tables 5.3, 5.4 and 5.5 demonstrate the importance of relating specific processes with changes taking place at appropriate points along the score distribution. These proficiency probabilities were defined in more detail in Chapter 4. Inspection of Table 5.4 indicates that gains between 8th and 12th grade in the probability of being proficient at level four (GPL4) show a positive correlation with number of units of mathematics of .44. The correlations between gains in probability of mastery and various course exposures vary some by estimation method, but in general the one-step Bayesian procedure does as well as the other methods. One of the primary purposes of the proficiency levels is to provide information for each individual on where on the scale his or her changes are taking place. For example, an individual who had a high scale score (on the Theta or "true score scale) in tenth grade and then received an even higher score in the twelfth grade would show his or her greatest gains in probability of mastery at either levels 4 or 5, the levels that mark the upper end of the scale.

When the "dummy" variable contrasting whether an individual is in the academic curriculum, coded "1" versus the general/vocational curriculum coded "0" is correlated with gains in probabilities at the various proficiency levels, one observes negative correlations for demonstrated proficiencies at the two lower levels (simple operations and fractions and decimals) and increasingly higher positive correlation for levels 3 through 5. That is, individuals with a score of "1" on the dummy variable indicating they are in the academic curriculum are making progressively greater gains in probabilities associated with mastery of levels 3 through 5. Conversely individuals who are coded "0" indicating that they are in the general/vocational curriculum are making their greatest gains in the two lower levels (simple operations and decimals/fractions). These general/vocational students' gains are typically taking place at the lower end of the scale and thus the negative correlation in the last column of Table 5.3. They are increasing their probabilities of proficiency primarily at the two lowest levels.

Tables 5.6-5.11 present similar correlations for reading, science, and H/C/G respectively. The ST1 procedure was selected on the basis of the math test results, so only ST1 estimates were computed for these content areas.

Table 5.6
Correlations of Background Variables with Second Follow-up Status and Gains Reading

	Gender	Ethnicity		SES		School Type		Curric
	Male=1 Female=0	Hisp=1 White=0	Black=1 White=0	Contin-uous	Quartile	Cath=1 Public=0	NAIS=1 Public=0	Acad=1 G+V=0
Second Follow-up Status								
IRT Number Right	-0.11	-0.16	-0.21	0.38	0.36	0.10	0.12	0.34
Standardized Theta	-0.11	-0.15	-0.21	0.38	0.36	0.10	0.12	0.34
Proficiency Level 1	-0.09	-0.06	-0.13	0.16	0.15	0.05	0.04	0.14
Proficiency Level 2	-0.10	-0.16	-0.21	0.35	0.34	0.10	0.09	0.31
Proficiency Level 3	-0.07	-0.12	-0.15	0.33	0.31	0.08	0.14	0.31
Gain: Base Year to First Follow-up								
IRT Number Right	-0.01	-0.03	-0.04	0.12	0.12	0.03	0.03	0.12
Standardized Theta	-0.01	-0.04	-0.05	0.14	0.13	0.03	0.05	0.14
Proficiency Level 1	0.01	0.03	0.02	-0.03	-0.03	-0.02	-0.01	-0.03
Proficiency Level 2	-0.02	-0.01	-0.03	0.07	0.07	0.03	-0.02	0.07
Proficiency Level 3	0.00	-0.09	-0.10	0.24	0.22	0.05	0.13	0.22
Gain: First to Second Follow-up								
IRT Number Right	-0.02	0.02	-0.01	-0.04	-0.03	0.01	-0.02	-0.04
Standardized Theta	-0.02	0.01	-0.02	-0.02	-0.01	0.02	-0.01	-0.02
Proficiency Level 1	0.00	0.04	0.02	-0.07	-0.07	-0.01	-0.01	-0.06
Proficiency Level 2	0.01	0.02	0.00	-0.08	-0.07	-0.01	-0.02	-0.08
Proficiency Level 3	-0.05	-0.03	-0.05	0.09	0.09	0.04	-0.01	0.09

Table 5.6 (cont'd)
Correlations of Background Variables with Second Follow-up Status and Gains
Reading

	Gender	Ethnicity		SES		School Type		Curric
	Male=1 Female=0	Hisp=1 White=0	Black=1 White=0	Contin-uous	Quartile	Cath=1 Public=0	NAIS=1 Public=0	Acad=1 G+V=0
Total Gain: Base Year to Second Follow-up								
IRT Number Right	-0.02	-0.01	-0.05	0.07	0.07	0.03	0.01	0.07
Standardized Theta	-0.03	-0.02	-0.06	0.10	0.10	0.04	0.03	0.10
Proficiency Level 1	0.01	0.06	0.04	-0.09	-0.08	-0.02	-0.02	-0.08
Proficiency Level 2	-0.01	0.01	-0.03	0.00	0.01	0.02	-0.03	0.00
Proficiency Level 3	-0.05	-0.10	-0.13	0.28	0.26	0.08	0.11	0.26

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 5.7
Correlations of Transcript Variables
with Second Follow-up Status and Gains
Reading

	Total # Units	Average Grades
Second Follow-up Status		
IRT Number Right	0.26	0.52
Standardized Theta	0.26	0.53
Proficiency Level 1	0.16	0.22
Proficiency Level 2	0.25	0.49
Proficiency Level 3	0.17	0.45
Gain: Base Year to First Follow-up		
IRT Number Right	0.13	0.16
Standardized Theta	0.13	0.18
Proficiency Level 1	0.00	-0.06
Proficiency Level 2	0.11	0.10
Proficiency Level 3	0.12	0.30
Gain: First to Second Follow-up		
IRT Number Right	0.00	-0.01
Standardized Theta	0.00	0.02
Proficiency Level 1	-0.06	-0.07
Proficiency Level 2	0.00	-0.06
Proficiency Level 3	0.06	0.14
Total Gain: Base Year to Second Follow-up		
IRT Number Right	0.11	0.13
Standardized Theta	0.12	0.18
Proficiency Level 1	-0.05	-0.11
Proficiency Level 2	0.09	0.03
Proficiency Level 3	0.16	0.38

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 5.8
Correlations of Background Variables with Second Follow-up Status and Gains
Science

	Gender	Ethnicity		SES		School Type		Curric	Taking Science	
	Male=1 Female=0	Hisp=1 White=0	Black=1 White=0	Contin-uous	Quartile	Cath=1 Public=0	NAIS=1 Public=0	Acad=1 G+V=0	Last 2 Years	Now
Second Follow-up Status										
IRT Number Right	0.16	-0.20	-0.30	0.41	0.39	0.07	0.11	0.35	0.21	0.31
Standardized Theta	0.16	-0.20	-0.30	0.41	0.38	0.07	0.11	0.35	0.21	0.31
Proficiency Level 1	0.07	-0.15	-0.27	0.26	0.24	0.05	0.05	0.21	0.15	0.16
Proficiency Level 2	0.15	-0.20	-0.28	0.38	0.37	0.07	0.10	0.33	0.20	0.29
Proficiency Level 3	0.15	-0.13	-0.18	0.34	0.32	0.05	0.11	0.30	0.17	0.29
Gain: Base Year to First Follow-up										
IRT Number Right	0.08	-0.08	-0.12	0.19	0.19	0.02	0.06	0.13	0.07	0.11
Standardized Theta	0.08	-0.07	-0.11	0.18	0.17	0.01	0.06	0.12	0.06	0.10
Proficiency Level 1	-0.01	0.04	0.04	-0.04	-0.04	-0.01	-0.02	-0.06	-0.03	-0.06
Proficiency Level 2	0.06	-0.09	-0.13	0.18	0.18	0.02	0.03	0.13	0.06	0.10
Proficiency Level 3	0.12	-0.09	-0.13	0.26	0.24	0.02	0.13	0.21	0.11	0.20
Gain: First to Second Follow-up										
IRT Number Right	0.05	0.01	-0.04	0.00	0.00	0.04	-0.03	0.03	0.06	0.06
Standardized Theta	0.06	0.00	-0.04	0.01	0.00	0.04	-0.02	0.03	0.06	0.07
Proficiency Level 1	0.02	0.04	0.03	-0.08	-0.09	-0.01	-0.01	-0.07	-0.01	-0.03
Proficiency Level 2	0.03	0.01	-0.03	-0.02	-0.02	0.04	-0.03	0.01	0.05	0.02
Proficiency Level 3	0.05	-0.05	-0.08	0.13	0.13	0.05	-0.02	0.13	0.08	0.13

Table 5.8 (cont'd)
Correlations of Background Variables with Second Follow-up Status and Gains
Science

	Gender	Ethnicity		SES		School Type		Curric	Taking Science	
	Male=1 Female=0	Hisp=1 White=0	Black=1 White=0	Contin-uous	Quartile	Cath=1 Public=0	NAIS=1 Public=0	Acad=1 G+V=0	Last 2 Years	Now
Total Gain: Base Year to Second Follow-up										
IRT Number Right	0.12	-0.07	-0.15	0.18	0.17	0.05	0.03	0.15	0.11	0.16
Standardized Theta	0.13	-0.06	-0.14	0.18	0.16	0.04	0.04	0.14	0.11	0.16
Proficiency Level 1	0.01	0.07	0.06	-0.11	-0.11	-0.02	-0.03	-0.11	-0.03	-0.08
Proficiency Level 2	0.08	-0.07	-0.15	0.14	0.15	0.05	0.00	0.12	0.09	0.11
Proficiency Level 3	0.13	-0.12	-0.17	0.31	0.29	0.06	0.09	0.27	0.16	0.26

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 5.9
Correlations of Transcript Variables
with Second Follow-up Status and Gains
Science

	Number of Units						
	Total # Units	Average Grade	Earth Science	Biology	Chemis-try	Physics	Other
Second Follow-up Status							
IRT Number Right	0.44	0.48	0.02	0.22	0.43	0.43	-0.16
Standardized Thet	0.43	0.48	0.01	0.22	0.43	0.43	-0.16
Proficiency Level 1	0.25	0.27	0.03	0.16	0.25	0.20	-0.09
Proficiency Level 2	0.41	0.45	0.02	0.22	0.41	0.39	-0.15
Proficiency Level 3	0.38	0.44	-0.01	0.15	0.38	0.43	-0.13
Gain: Base Year to First Follow-up							
IRT Number Right	0.21	0.21	0.02	0.10	0.18	0.19	-0.04
Standardized Theta	0.20	0.19	0.02	0.08	0.16	0.18	-0.03
Proficiency Level 1	-0.04	-0.10	0.03	-0.02	-0.05	-0.07	0.02
Proficiency Level 2	0.20	0.21	0.02	0.11	0.17	0.14	-0.03
Proficiency Level 3	0.28	0.32	-0.03	0.10	0.25	0.36	-0.09
Gain: First to Second Follow-up							
IRT Number Right	0.01	-0.01	0.00	0.00	0.02	0.00	0.00
Standardized Theta	0.01	0.00	0.00	-0.01	0.02	0.01	0.00
Proficiency Level 1	-0.09	-0.10	-0.01	-0.05	-0.11	-0.07	0.06
Proficiency Level 2	-0.02	-0.05	0.00	0.00	0.01	-0.03	-0.01
Proficiency Level 3	0.15	0.16	0.01	0.06	0.17	0.11	-0.04
Total Gain: Base Year to Second Follow-up							
IRT Number Right	0.21	0.20	0.02	0.09	0.19	0.19	-0.04
Standardized Theta	0.20	0.19	0.02	0.07	0.17	0.19	-0.04
Proficiency Level 1	-0.12	-0.18	0.02	-0.07	-0.14	-0.13	0.07
Proficiency Level 2	0.17	0.15	0.02	0.11	0.17	0.11	-0.04
Proficiency Level 3	0.35	0.39	-0.01	0.14	0.34	0.38	-0.11

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Table 5.10
Correlations of Background Variables with Second Follow-up Status and Gains History/Citizenship/Geography

	Gender	Ethnicity		SES		School Type		Curric
	Male=1 Female=0	Hisp=1 White=0	Black=1 White=0	Contin-uous	Quartile	Cath=1 Public=0	NAIS=1 Public=0	Acad=1 G+V=0
Second Follow-up Status								
IRT Number Right	0.08	-0.15	-0.20	0.41	0.39	0.11	0.11	0.36
Standardized Theta	0.08	-0.15	-0.20	0.41	0.38	0.11	0.12	0.36
Gain: Base Year to First Follow-up								
IRT Number Right	0.02	-0.01	-0.03	0.08	0.09	0.00	-0.03	0.09
Standardized Theta	0.01	0.00	-0.02	0.06	0.06	-0.01	-0.03	0.08
Gain: First to Second Follow-up								
IRT Number Right	0.01	0.01	-0.03	0.05	0.05	0.06	0.02	0.03
Standardized Theta	0.02	0.01	-0.02	0.04	0.04	0.06	0.03	0.02
Total Gain: Base Year to Second Follow-up								
IRT Number Right	0.03	-0.01	-0.06	0.12	0.12	0.05	-0.01	0.11
Standardized Theta	0.02	0.01	-0.04	0.09	0.09	0.05	0.00	0.09

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

98

Table 5.11
Correlations of Transcript Variables with Second Follow-up Status and Gains History/Citizenship/Geography

	Number of Units			
	Total # Units	Average Grade	History	Other
Second Follow-up Status				
IRT Number Right	0.25	0.55	0.24	0.11
Standardized Theta	0.25	0.54	0.24	0.11
Gain: Base Year to First Follow-up				
IRT Number Right	0.11	0.14	0.08	0.06
Standardized Theta	0.09	0.10	0.06	0.06
Gain: First to Second Follow-up				
IRT Number Right	0.02	0.07	0.00	0.03
Standardized Theta	0.01	0.06	-0.01	0.02
Total Gain: Base Year to Second Follow-up				
IRT Number Right	0.11	0.18	0.06	0.08
Standardized Theta	0.09	0.14	0.05	0.07

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

The reader should note that the column labeled "total units" refers to the total number of semesters of mathematics, english, science or social studies courses taken depending on the content area being analyzed. As in the case of mathematics, the pattern of the total score gains and the proficiency probability gains were consistent with our theoretical expectations. That is, the aggregate (total) score gains show the expected patterns of overall gain while gains in proficiency probabilities show maximum relationships with school process that target learning that is appropriate for that particular mastery level.

References

Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. *Journal of Educational Measurement, 20,* 103-118.

Baker, F. B. (1992). *Item Response Theory.* New York: Marcel-Dekker, Inc.

Bock, R. D. (Ed.) (1989). *Multilevel Analysis of Educational Data.* San Diego: Academic.

Bock, R. D., Muraki, E. J. & Pfeiffenberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement, 25,* 275-285.

Cole, N. S., & Moss, P. A. (1989). Bias in test use. In Robert L. Linn (Ed.)

Cole, N. S., & Nitko, A. J. (1981). Instrumentation and bias: Issues in selecting measures for educational evaluations. In R. A. Berk (Ed.), *Educational Evaluation Methodology: The State of the Art.* Baltimore: Johns Hopkins University Press.

Cooley, W. W. (1977, August). *Program evaluation in education.* Invited paper presented at the meeting of the American Psychological Association, San Francisco.

Cronbach, L. J. (1963). Evaluation for course improvement. *Teachers College Record, 64,* 672-683.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal Royal Statistical Society,* Series B. *39,* 1-38.

Frechtling, J. A. (1989). Administrative uses of school testing programs. In R. L. Linn (ed.) *Educational Measurement 3rd Edit.,* 475-483. New York: MacMillan.

Holland, P. W. & Thayer, D. T. (1986). *Differential Item function and the Mantel-Haenszel Procedure.* ETS Research Report No. 86-31. Princeton, NJ.

Ingels, S.J., et al. (1987). *Field Test Report: National Education Longitudinal Study of 1988 (Base Year).* (Chapter 3: Rock and Pollack, The Cognitive Test Battery). Chicago: NORC, University of Chicago. (ERIC ED 289-897).

Ingels, S.J., Dowd, K.L., Baldridge, J.D., Stipe, J.L., Bartot, V.H., and Frankel, M.R. (1994). *NELS:88 Second Follow-Up Student Component Data File User's Manual.* NCES 93-374, Washington, D.C.: National Center for Education Statistics.

Kirsch, I.S., et al. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey.* Washington, DC: National Center for Education Statistics.

Leinhardt, G. (1980). Modeling and measuring educational treatment in evaluation. *Review of Educational Research, 50,* 393-420.

Leinhardt, G., & Seewald, A. M. (1981). Overlap: What's tested, what's taught? *Journal of Educational Measurement, 18,* 85-96.

Linn, R. L., & Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement, 18,* 109-118.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, N. J.: Erlbaum.

Mantel, & Haenszel (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Mehrens, W. A. (1984). National tests and local curriculum: Match or mismatch? *Educational Measurement: Issues and Practice, 3* (3), 9-15.

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. *Journal of Educational Measurement, 23,* 185-196.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association, 80,* 993-997.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49,* 359-381.

Mislevy, R. J., et al. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics, 17,* 131-154.

Mislevy R. J. & Bock, R. D. (1982). *BILOG: Item analysis and test scoring with binary logistic models.* [Computer program]. Mooresville, IN: Scientific Software.

Mislevy, R. J., & Bock, R. D. (1989). *PC-BILOG 3: Item analysis and test scoring with binary logistic models.* Mooresville, IN: Scientific Software, Inc.

Muraki, E. J. & Bock, R. D. (1987). *BIMAIN: A program for item pool maintenance in the presence of item parameter drift and item bias.* Mooresville, IN: Scientific Software, Inc.

Muraki, E. J. & Bock, R. D. (1991). *PARSCALE: Parameter scaling of rating data* [Computer program]. Chicago. IL: Scientific Software, Inc.

Nitko, A. J. (1983). *Educational tests and measurements: An introduction.* San Diego: Harcourt, Brace Jovanovich.

Nitko, A. J. (1989). Designing tests that are integrated with instruction. In Robert L. Linn (Ed.) *Educational Measurement,* 3rd edition. National Council on Measurement in Education and American Council on Education. New York: MacMillan Publishing Company.

Phillips, S. E., & Mehrens, W. A. (1988). Effects of curricular differences on achievement test data at item and objective levels. *Applied Measurement in Education, 1,* 33-51.

Rasinski, K.A., Ingels, S.J., Rock, D.A., & Pollack, J.M. (1993). *America's High School Sophomores: A Ten Year Comparison.* NCES 93-087, Washington, D.C.: National Center for Education Statistics.

Rock, D.A., Hilton, T.L., Pollack, J.M., Ekstrom, R.B., & Goertz, M.E. (1985). *Psychometric Analysis of the NLS-72 and the High School and Beyond Test Batteries*. NCES 85-217, Washington, D.C.: National Center for Education Statistics.

Rock, D.A. & Pollack, J. (1987). The Cognitive Test Battery. In S.J. Ingels et al., *Field Test Report: National Education Longitudinal Study of 1988 (Base Year)*. Chicago: NORC, University of Chicago. (ERIC ED 289-897).

Rock, D.A., Owings, J. & Lee, R. (1994). *Changes in Math Proficiency between Eighth and Tenth Grades*. (NCES 93-455). *Statistics in Brief*, January 1994.

Rock, D. A., & Pollack, J. M. (1991). *Psychometric Report for the NELS:88 Base Year Test Battery* (NCES Report No. 91-468). Washington, DC: National Center for Education Statistics.

Samejima, (1994). Roles of Fisher type information in latent trait models. In H. Bozdogan (Ed.) *Frontiers of Statistical Modeling:* Netherlands, Kluwer Publishing

Schmidt, W. H. (1983). Content bias in achievement tests. *Journal of Educational Measurement, 20,* 165-178.

Scott, L.A., Rock, D.A., Pollack, J.M., & Ingels, S.J. (1995). *Two Years Later: Cognitive Gains and School Transitions of NELS:88 Eighth Graders*. NCES Report 95-436. Washington, D.C.: National Center for Education Statistics.

Stocking, M. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST: User's Guide*. Princeton, NJ: Educational Testing Service.

Yamamoto, K. & Mazzeo, J. (1992). Item response theory: Scale linking in NAEP. *Journal of Educational Statistics, 17,* pp. 155-173.

Appendix A: Reading

Appendix A-1
Reading: Base Year (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.95	0.93	0.96	0.95	0.93	0.93	0.95	0.60	0.61	0.57	0.68	0.54	0.51	0.64
Item 2	0.85	0.85	0.86	0.85	0.79	0.75	0.88	0.63	0.62	0.64	0.67	0.59	0.55	0.62
Item 3	0.82	0.80	0.85	0.80	0.75	0.73	0.85	0.65	0.63	0.67	0.72	0.62	0.58	0.65
Item 4	0.57	0.53	0.62	0.56	0.46	0.38	0.63	0.67	0.66	0.67	0.64	0.64	0.63	0.65
Item 5	0.55	0.53	0.57	0.55	0.41	0.44	0.59	0.67	0.63	0.71	0.69	0.63	0.60	0.67
Item 15	0.60	0.61	0.60	0.62	0.49	0.44	0.65	0.65	0.68	0.63	0.72	0.61	0.56	0.64
Item 16	0.41	0.39	0.42	0.43	0.29	0.26	0.45	0.63	0.64	0.62	0.69	0.56	0.52	0.62
Item 17	0.49	0.48	0.50	0.54	0.37	0.36	0.54	0.68	0.66	0.69	0.71	0.66	0.61	0.67
Item 18	0.61	0.56	0.66	0.66	0.55	0.51	0.64	0.57	0.55	0.57	0.52	0.54	0.54	0.57
Item 19	0.39	0.38	0.41	0.43	0.34	0.32	0.42	0.44	0.50	0.39	0.45	0.45	0.39	0.44
Item 20	0.59	0.54	0.63	0.64	0.54	0.46	0.62	0.64	0.65	0.62	0.64	0.54	0.55	0.66
Item 22	0.71	0.66	0.76	0.70	0.61	0.52	0.76	0.75	0.75	0.75	0.77	0.67	0.66	0.76
Item 23	0.50	0.52	0.49	0.54	0.43	0.39	0.54	0.55	0.56	0.55	0.61	0.43	0.37	0.57
Item 24	0.48	0.45	0.50	0.52	0.38	0.37	0.51	0.65	0.64	0.65	0.70	0.52	0.53	0.66
Item 38	0.47	0.43	0.49	0.51	0.36	0.36	0.50	0.70	0.70	0.69	0.73	0.64	0.69	0.69
Item 39	0.76	0.74	0.79	0.79	0.67	0.65	0.80	0.74	0.74	0.72	0.71	0.66	0.65	0.75
Item 40	0.54	0.50	0.58	0.57	0.39	0.40	0.58	0.66	0.63	0.69	0.65	0.53	0.49	0.69
Item 41	0.54	0.51	0.56	0.56	0.48	0.45	0.56	0.53	0.51	0.55	0.51	0.46	0.51	0.53
Item 42	0.63	0.60	0.66	0.65	0.52	0.45	0.67	0.67	0.65	0.69	0.69	0.56	0.58	0.68
Item 43	0.70	0.67	0.74	0.73	0.63	0.57	0.74	0.64	0.63	0.65	0.63	0.57	0.55	0.65
Item 44	0.62	0.60	0.64	0.63	0.50	0.48	0.67	0.62	0.59	0.65	0.67	0.53	0.48	0.63
Mean	0.61	0.58	0.63	0.63	0.52	0.49	0.65	0.63	0.63	0.64	0.66	0.57	0.55	0.64
S.D.	0.14	0.14	0.15	0.13	0.16	0.16	0.14	0.07	0.06	0.08	0.08	0.07	0.08	0.07
N	23643	11755	11888	1507	3007	2878	15849							
Wtd N	2897540	1451017	1446523	102799	293439	376518	2072285							
% Answering Last Item	96%	95%	96%	96%	93%	90%	97%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

BEST COPY AVAILABLE

Appendix A-2
Reading: First Follow-up (Low Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.92	0.91	0.93	0.94	0.94	0.91	0.92	0.63	0.66	0.58	0.59	0.60	0.54	0.67
Item 2	0.80	0.80	0.80	0.79	0.79	0.75	0.82	0.61	0.64	0.59	0.62	0.61	0.60	0.62
Item 3	0.77	0.75	0.79	0.75	0.75	0.72	0.79	0.65	0.63	0.66	0.69	0.64	0.65	0.64
Item 4	0.50	0.49	0.52	0.48	0.46	0.38	0.56	0.59	0.57	0.61	0.50	0.59	0.56	0.58
Item 5	0.46	0.44	0.48	0.44	0.40	0.45	0.48	0.58	0.54	0.63	0.66	0.57	0.54	0.59
Item 15	0.54	0.56	0.51	0.56	0.52	0.42	0.58	0.61	0.65	0.58	0.54	0.62	0.51	0.62
Item 16	0.33	0.34	0.32	0.31	0.27	0.26	0.37	0.51	0.53	0.50	0.59	0.49	0.36	0.52
Item 17	0.44	0.45	0.42	0.50	0.40	0.32	0.48	0.61	0.63	0.60	0.55	0.63	0.54	0.62
Item 18	0.54	0.51	0.58	0.57	0.59	0.49	0.55	0.45	0.45	0.44	0.43	0.48	0.54	0.42
Item 19	0.36	0.36	0.36	0.43	0.35	0.33	0.37	0.41	0.42	0.40	0.46	0.41	0.41	0.40
Item 21	0.65	0.58	0.72	0.68	0.63	0.64	0.65	0.59	0.58	0.59	0.69	0.54	0.62	0.59
Item 22	0.62	0.55	0.70	0.66	0.60	0.54	0.66	0.69	0.67	0.70	0.70	0.61	0.66	0.70
Item 23	0.48	0.49	0.46	0.46	0.45	0.44	0.49	0.48	0.48	0.49	0.34	0.45	0.37	0.52
Item 24	0.41	0.39	0.43	0.42	0.38	0.39	0.43	0.58	0.59	0.56	0.65	0.44	0.59	0.60
Item 38	0.38	0.37	0.40	0.39	0.41	0.34	0.39	0.61	0.65	0.56	0.59	0.61	0.58	0.62
Item 39	0.71	0.68	0.74	0.77	0.68	0.67	0.73	0.72	0.73	0.69	0.66	0.66	0.70	0.74
Item 40	0.40	0.38	0.41	0.45	0.33	0.35	0.43	0.52	0.51	0.52	0.49	0.43	0.35	0.57
Item 41	0.46	0.46	0.47	0.53	0.46	0.45	0.47	0.47	0.46	0.47	0.47	0.47	0.49	0.46
Item 42	0.55	0.53	0.57	0.61	0.53	0.46	0.58	0.64	0.63	0.64	0.67	0.58	0.57	0.66
Item 43	0.67	0.63	0.72	0.71	0.66	0.61	0.69	0.58	0.57	0.57	0.62	0.46	0.57	0.60
Item 44	0.55	0.54	0.57	0.49	0.51	0.48	0.59	0.53	0.53	0.53	0.51	0.46	0.42	0.56
Mean	0.55	0.53	0.57	0.57	0.53	0.50	0.57	0.57	0.58	0.57	0.57	0.54	0.53	0.59
S.D.	0.15	0.15	0.16	0.15	0.16	0.16	0.15	0.08	0.08	0.08	0.10	0.08	0.10	0.08
N	9115	4890	4225	531	1542	1369	5457							
Wtd N	1511539	818585	692954	55097	207601	285743	924408							
% Answering Last Item	94%	95%	94%	92%	89%	90%	97%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix A-3
Reading: First Follow-up (High Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 6	0.63	0.64	0.61	0.66	0.52	0.65	0.63	0.51	0.49	0.53	0.62	0.58	0.48	0.49
Item 7	0.55	0.56	0.54	0.59	0.49	0.46	0.56	0.53	0.52	0.55	0.53	0.46	0.47	0.54
Item 8	0.55	0.59	0.52	0.56	0.46	0.55	0.56	0.57	0.62	0.54	0.64	0.57	0.49	0.57
Item 9	0.66	0.65	0.66	0.70	0.60	0.66	0.66	0.70	0.73	0.68	0.59	0.67	0.68	0.71
Item 10	0.57	0.54	0.60	0.59	0.47	0.59	0.58	0.53	0.50	0.56	0.67	0.52	0.41	0.54
Item 11	0.84	0.82	0.86	0.84	0.79	0.84	0.85	0.72	0.76	0.67	0.58	0.80	0.74	0.71
Item 12	0.60	0.61	0.59	0.64	0.51	0.54	0.61	0.62	0.66	0.59	0.56	0.63	0.49	0.63
Item 13	0.76	0.76	0.76	0.77	0.62	0.64	0.78	0.70	0.69	0.71	0.65	0.72	0.67	0.70
Item 15	0.86	0.88	0.85	0.87	0.81	0.84	0.87	0.68	0.74	0.65	0.62	0.67	0.72	0.68
Item 16	0.67	0.68	0.66	0.73	0.63	0.52	0.69	0.61	0.64	0.59	0.57	0.63	0.44	0.62
Item 17	0.81	0.80	0.83	0.83	0.79	0.80	0.82	0.69	0.76	0.62	0.71	0.59	0.71	0.70
Item 19	0.52	0.53	0.51	0.51	0.46	0.42	0.53	0.41	0.40	0.43	0.38	0.41	0.40	0.41
Item 20	0.76	0.72	0.80	0.81	0.73	0.76	0.77	0.59	0.60	0.58	0.62	0.56	0.57	0.60
Item 22	0.91	0.89	0.93	0.93	0.87	0.86	0.92	0.75	0.75	0.75	0.96	0.82	0.60	0.75
Item 23	0.79	0.79	0.79	0.81	0.67	0.73	0.81	0.66	0.65	0.67	0.63	0.52	0.50	0.68
Item 24	0.82	0.79	0.84	0.82	0.75	0.80	0.83	0.73	0.72	0.74	0.79	0.65	0.64	0.74
Item 29	0.51	0.50	0.52	0.56	0.43	0.47	0.52	0.50	0.46	0.53	0.60	0.52	0.40	0.49
Item 30	0.63	0.65	0.62	0.72	0.60	0.64	0.63	0.47	0.55	0.41	0.43	0.32	0.30	0.50
Item 31	0.78	0.74	0.81	0.82	0.70	0.74	0.79	0.65	0.68	0.62	0.63	0.67	0.54	0.65
Item 32	0.45	0.45	0.46	0.52	0.41	0.39	0.46	0.48	0.48	0.49	0.50	0.49	0.52	0.48
Item 33	0.36	0.36	0.36	0.49	0.33	0.40	0.35	0.41	0.42	0.41	0.34	0.44	0.43	0.41
Mean	0.67	0.66	0.67	0.70	0.60	0.63	0.68	0.60	0.61	0.59	0.60	0.58	0.53	0.60
S.D.	0.15	0.14	0.15	0.13	0.15	0.15	0.15	0.10	0.12	0.10	0.13	0.12	0.12	0.10
N	8717	4023	4694	587	668	491	6914							
Wtd N	1368601	629586	739015	50541	84488	88657	1135773							
% Answering Last Item	98%	98%	98%	97%	96%	93%	98%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix A-4
Reading: Second Follow-up (Low Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.93	0.92	0.95	0.93	0.94	0.92	0.94	0.64	0.66	0.60	0.62	0.64	0.62	0.68
Item 2	0.82	0.82	0.82	0.84	0.80	0.76	0.84	0.66	0.65	0.67	0.65	0.62	0.64	0.66
Item 3	0.80	0.77	0.84	0.78	0.78	0.78	0.82	0.67	0.68	0.65	0.76	0.60	0.62	0.68
Item 4	0.57	0.54	0.60	0.48	0.55	0.44	0.62	0.64	0.65	0.63	0.66	0.53	0.67	0.65
Item 5	0.56	0.55	0.58	0.48	0.45	0.54	0.61	0.62	0.61	0.64	0.68	0.64	0.64	0.60
Item 14	0.25	0.29	0.19	0.24	0.20	0.18	0.28	0.47	0.52	0.43	0.27	0.46	0.37	0.49
Item 15	0.58	0.59	0.57	0.58	0.55	0.45	0.63	0.70	0.72	0.67	0.65	0.72	0.68	0.68
Item 16	0.36	0.37	0.35	0.35	0.31	0.30	0.40	0.61	0.65	0.57	0.68	0.65	0.62	0.58
Item 17	0.45	0.46	0.44	0.50	0.40	0.37	0.49	0.62	0.63	0.61	0.66	0.64	0.54	0.62
Item 19	0.36	0.36	0.36	0.39	0.35	0.29	0.39	0.37	0.38	0.36	0.31	0.41	0.36	0.35
Item 22	0.63	0.58	0.70	0.62	0.58	0.51	0.69	0.69	0.70	0.66	0.65	0.65	0.68	0.68
Item 23	0.53	0.52	0.53	0.54	0.48	0.44	0.57	0.52	0.55	0.49	0.54	0.40	0.48	0.54
Item 24	0.47	0.44	0.51	0.51	0.41	0.42	0.50	0.62	0.61	0.63	0.73	0.59	0.60	0.62
Item 38	0.48	0.48	0.49	0.50	0.51	0.46	0.49	0.66	0.68	0.64	0.70	0.68	0.62	0.67
Item 39	0.79	0.77	0.81	0.80	0.77	0.75	0.80	0.69	0.69	0.68	0.64	0.64	0.66	0.71
Item 41	0.54	0.54	0.55	0.63	0.54	0.54	0.54	0.50	0.48	0.53	0.36	0.42	0.54	0.53
Item 45	0.64	0.65	0.63	0.67	0.63	0.56	0.66	0.53	0.51	0.56	0.64	0.54	0.44	0.54
Item 46	0.42	0.41	0.43	0.42	0.43	0.37	0.42	0.33	0.34	0.32	0.39	0.27	0.24	0.36
Item 47	0.68	0.62	0.76	0.79	0.74	0.61	0.69	0.59	0.58	0.59	0.63	0.52	0.51	0.62
Item 48	0.35	0.33	0.37	0.44	0.33	0.30	0.36	0.45	0.44	0.46	0.51	0.48	0.39	0.44
Item 49	0.34	0.34	0.34	0.37	0.33	0.30	0.35	0.39	0.41	0.36	0.38	0.41	0.40	0.38
Mean	0.55	0.54	0.56	0.56	0.53	0.49	0.58	0.57	0.58	0.56	0.58	0.55	0.54	0.58
S.D.	0.18	0.17	0.19	0.18	0.19	0.19	0.18	0.11	0.11	0.11	0.14	0.12	0.13	0.11
N	7076	3808	3268	450	1209	1008	4258							
Wtd N	1222645	675058	547587	49551	171255	216162	757448							
% Answering Last Item	93%	93%	93%	87%	87%	90%	95%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix A-5
Reading: Second Follow-up (High Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 19	0.57	0.58	0.56	0.51	0.47	0.46	0.59	0.43	0.42	0.45	0.44	0.39	0.34	0.43
Item 22	0.94	0.92	0.96	0.94	0.93	0.88	0.95	0.66	0.72	0.57	0.55	0.65	0.50	0.68
Item 23	0.86	0.84	0.87	0.89	0.79	0.74	0.87	0.61	0.62	0.59	0.48	0.59	0.35	0.64
Item 24	0.89	0.86	0.91	0.94	0.83	0.84	0.89	0.65	0.67	0.61	0.61	0.65	0.54	0.65
Item 25	0.47	0.50	0.44	0.49	0.42	0.36	0.48	0.46	0.46	0.47	0.54	0.39	0.30	0.47
Item 26	0.70	0.68	0.72	0.68	0.63	0.61	0.71	0.47	0.46	0.47	0.43	0.55	0.48	0.46
Item 27	0.90	0.88	0.91	0.90	0.82	0.79	0.91	0.45	0.47	0.42	0.39	0.37	0.46	0.44
Item 28	0.87	0.86	0.88	0.90	0.84	0.81	0.88	0.62	0.62	0.62	0.53	0.56	0.45	0.64
Item 34	0.59	0.58	0.60	0.74	0.51	0.45	0.60	0.51	0.57	0.45	0.50	0.54	0.57	0.49
Item 35	0.32	0.34	0.30	0.41	0.24	0.28	0.32	0.47	0.49	0.47	0.68	0.42	0.37	0.47
Item 36	0.50	0.47	0.53	0.61	0.50	0.40	0.51	0.59	0.61	0.58	0.74	0.59	0.59	0.59
Item 37	0.42	0.38	0.46	0.48	0.34	0.34	0.43	0.55	0.50	0.59	0.69	0.57	0.46	0.55
Item 45	0.84	0.83	0.85	0.88	0.79	0.72	0.85	0.66	0.64	0.68	0.62	0.57	0.54	0.67
Item 46	0.61	0.62	0.60	0.68	0.54	0.48	0.62	0.61	0.60	0.63	0.64	0.68	0.54	0.60
Item 48	0.52	0.51	0.53	0.61	0.42	0.46	0.53	0.54	0.54	0.55	0.59	0.55	0.47	0.54
Item 49	0.56	0.55	0.57	0.66	0.45	0.44	0.57	0.60	0.55	0.64	0.49	0.57	0.66	0.59
Item 50	0.77	0.77	0.78	0.86	0.71	0.70	0.78	0.60	0.63	0.57	0.54	0.58	0.54	0.61
Item 51	0.49	0.45	0.52	0.56	0.49	0.51	0.49	0.47	0.46	0.47	0.62	0.47	0.34	0.48
Item 52	0.43	0.42	0.44	0.45	0.39	0.35	0.44	0.47	0.44	0.51	0.50	0.42	0.48	0.47
Item 53	0.44	0.44	0.44	0.51	0.29	0.37	0.45	0.44	0.45	0.44	0.51	0.43	0.32	0.44
Item 54	0.30	0.28	0.32	0.36	0.25	0.28	0.30	0.45	0.47	0.44	0.49	0.41	0.40	0.46
Mean	0.62	0.61	0.63	0.67	0.55	0.54	0.63	0.54	0.54	0.53	0.55	0.52	0.46	0.54
S.D.	0.20	0.20	0.20	0.19	0.21	0.19	0.20	0.08	0.09	0.08	0.09	0.09	0.10	0.08
N	7154	3311	3843	492	549	398	5671							
Wtd N	1058046	493754	564292	41193	64824	73813	872234							
% Answering Last Item	91%	91%	91%	92%	83%	75%	93%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix B: Math

Appendix B-1
Math: Base Year (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.56	0.54	0.59	0.60	0.43	0.37	0.62	0.60	0.62	0.57	0.60	0.54	0.48	0.58
Item 3	0.69	0.73	0.65	0.69	0.57	0.53	0.74	0.56	0.59	0.53	0.57	0.54	0.45	0.54
Item 5	0.52	0.50	0.53	0.60	0.43	0.36	0.55	0.66	0.68	0.65	0.76	0.60	0.60	0.66
Item 6	0.59	0.61	0.57	0.68	0.46	0.37	0.65	0.68	0.68	0.68	0.73	0.62	0.57	0.66
Item 7	0.65	0.71	0.59	0.70	0.53	0.34	0.73	0.65	0.63	0.69	0.66	0.61	0.56	0.61
Item 8	0.51	0.49	0.52	0.64	0.42	0.38	0.54	0.60	0.60	0.61	0.70	0.48	0.53	0.61
Item 9	0.62	0.63	0.61	0.65	0.55	0.49	0.66	0.60	0.60	0.60	0.72	0.49	0.49	0.61
Item 10	0.66	0.65	0.66	0.71	0.59	0.52	0.70	0.55	0.60	0.51	0.68	0.49	0.41	0.56
Item 11	0.51	0.52	0.50	0.61	0.41	0.36	0.55	0.65	0.65	0.65	0.79	0.57	0.56	0.65
Item 12	0.49	0.48	0.49	0.58	0.36	0.36	0.53	0.65	0.68	0.61	0.72	0.57	0.57	0.65
Item 13	0.44	0.46	0.42	0.55	0.36	0.35	0.46	0.51	0.50	0.52	0.63	0.34	0.32	0.55
Item 15	0.41	0.41	0.42	0.50	0.28	0.27	0.45	0.69	0.68	0.69	0.76	0.63	0.62	0.68
Item 16	0.44	0.42	0.46	0.55	0.35	0.32	0.47	0.66	0.66	0.67	0.71	0.63	0.60	0.65
Item 17	0.50	0.51	0.49	0.56	0.38	0.31	0.55	0.59	0.59	0.60	0.64	0.53	0.51	0.57
Item 18	0.47	0.46	0.49	0.51	0.42	0.44	0.49	0.27	0.28	0.26	0.37	0.22	0.23	0.27
Item 22	0.52	0.54	0.51	0.63	0.41	0.38	0.57	0.70	0.70	0.69	0.75	0.63	0.60	0.70
Item 23	0.41	0.40	0.41	0.42	0.28	0.26	0.46	0.60	0.63	0.56	0.62	0.46	0.49	0.59
Item 24	0.45	0.45	0.45	0.49	0.36	0.36	0.48	0.45	0.43	0.46	0.53	0.34	0.36	0.45
Item 25	0.37	0.36	0.37	0.53	0.28	0.23	0.40	0.58	0.56	0.61	0.66	0.44	0.46	0.59
Item 26	0.35	0.37	0.34	0.44	0.28	0.22	0.39	0.54	0.52	0.56	0.62	0.43	0.34	0.55
Item 28	0.50	0.52	0.48	0.61	0.38	0.37	0.54	0.69	0.68	0.69	0.72	0.65	0.60	0.69
Item 29	0.71	0.69	0.72	0.76	0.66	0.64	0.73	0.51	0.53	0.49	0.61	0.46	0.45	0.53
Item 30	0.79	0.79	0.79	0.82	0.75	0.73	0.80	0.50	0.55	0.44	0.57	0.49	0.44	0.50
Item 31	0.70	0.70	0.70	0.73	0.61	0.61	0.73	0.46	0.50	0.43	0.53	0.41	0.37	0.46
Item 32	0.52	0.51	0.52	0.63	0.42	0.38	0.55	0.64	0.63	0.65	0.75	0.50	0.54	0.65
Item 33	0.79	0.78	0.81	0.86	0.72	0.73	0.81	0.59	0.61	0.56	0.69	0.54	0.51	0.60
Item 34	0.46	0.48	0.45	0.50	0.41	0.36	0.49	0.31	0.32	0.29	0.40	0.33	0.28	0.28
Item 35	0.59	0.59	0.59	0.63	0.50	0.43	0.63	0.57	0.58	0.56	0.64	0.58	0.51	0.54
Item 36	0.52	0.50	0.54	0.57	0.44	0.42	0.55	0.54	0.57	0.52	0.62	0.50	0.51	0.54
Item 37	0.38	0.37	0.38	0.50	0.25	0.21	0.42	0.70	0.67	0.72	0.79	0.63	0.55	0.69
Item 38	0.45	0.46	0.43	0.46	0.29	0.20	0.52	0.70	0.70	0.69	0.70	0.70	0.63	0.66
Item 39	0.27	0.27	0.27	0.44	0.20	0.15	0.30	0.62	0.59	0.64	0.70	0.54	0.53	0.61
Item 40	0.41	0.43	0.39	0.45	0.37	0.35	0.43	0.32	0.35	0.29	0.42	0.29	0.24	0.32
Item 44	0.40	0.42	0.38	0.50	0.30	0.24	0.44	0.63	0.65	0.61	0.71	0.57	0.45	0.62
Item 50	0.56	0.58	0.54	0.60	0.46	0.43	0.60	0.50	0.53	0.47	0.50	0.41	0.34	0.50
Item 60	0.71	0.69	0.72	0.77	0.60	0.54	0.75	0.69	0.69	0.69	0.75	0.64	0.62	0.68
Item 61	0.79	0.76	0.82	0.86	0.75	0.72	0.81	0.51	0.56	0.46	0.56	0.53	0.46	0.50
Item 62	0.68	0.70	0.67	0.76	0.55	0.45	0.75	0.71	0.71	0.71	0.71	0.66	0.64	0.70
Item 63	0.65	0.66	0.65	0.68	0.56	0.53	0.69	0.45	0.43	0.46	0.37	0.46	0.49	0.41
Item 64	0.61	0.63	0.60	0.68	0.48	0.40	0.67	0.76	0.75	0.77	0.80	0.66	0.66	0.76
Mean	0.54	0.54	0.54	0.61	0.45	0.40	0.58	0.58	0.59	0.57	0.64	0.52	0.49	0.57
S.D.	0.13	0.13	0.13	0.11	0.14	0.14	0.13	0.11	0.11	0.12	0.11	0.11	0.11	0.11
N	23648	11763	11885	1503	3004	2885	15854							
Wtd N	2897116	1450776	1446340	102533	292817	376869	2072310							
% Answering Last Item	95%	95%	95%	96%	93%	90%	96%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix B-2
Math: First Follow-up (Low Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.42	0.41	0.43	0.55	0.41	0.38	0.44	0.41	0.41	0.41	0.37	0.38	0.42	0.39
Item 3	0.50	0.57	0.43	0.48	0.47	0.46	0.53	0.31	0.38	0.22	0.33	0.32	0.26	0.32
Item 4	0.83	0.82	0.84	0.80	0.84	0.81	0.85	0.49	0.54	0.46	0.52	0.44	0.50	0.50
Item 5	0.37	0.36	0.37	0.50	0.38	0.36	0.36	0.44	0.44	0.45	0.65	0.38	0.49	0.43
Item 6	0.45	0.49	0.42	0.47	0.42	0.36	0.52	0.49	0.56	0.42	0.62	0.50	0.39	0.51
Item 7	0.47	0.58	0.37	0.54	0.39	0.32	0.57	0.45	0.41	0.47	0.55	0.43	0.45	0.43
Item 8	0.44	0.44	0.43	0.61	0.43	0.46	0.42	0.46	0.41	0.52	0.56	0.41	0.43	0.50
Item 9	0.49	0.54	0.43	0.45	0.48	0.46	0.50	0.40	0.37	0.42	0.53	0.22	0.41	0.44
Item 10	0.51	0.51	0.51	0.45	0.52	0.51	0.51	0.38	0.40	0.36	0.53	0.33	0.35	0.42
Item 11	0.37	0.41	0.32	0.42	0.34	0.34	0.39	0.48	0.45	0.51	0.49	0.47	0.33	0.54
Item 12	0.35	0.32	0.38	0.48	0.35	0.33	0.36	0.41	0.48	0.37	0.62	0.43	0.41	0.40
Item 13	0.31	0.34	0.28	0.33	0.27	0.27	0.34	0.40	0.39	0.41	0.58	0.43	0.34	0.41
Item 14	0.71	0.73	0.69	0.77	0.71	0.73	0.70	0.51	0.52	0.50	0.58	0.42	0.46	0.57
Item 16	0.26	0.24	0.27	0.22	0.31	0.26	0.24	0.43	0.40	0.48	0.21	0.43	0.46	0.45
Item 19	0.27	0.23	0.30	0.37	0.31	0.22	0.27	0.36	0.36	0.38	0.53	0.36	0.31	0.36
Item 20	0.27	0.26	0.29	0.35	0.29	0.29	0.26	0.37	0.40	0.36	0.51	0.32	0.34	0.40
Item 21	0.54	0.51	0.56	0.46	0.57	0.51	0.55	0.40	0.40	0.40	0.50	0.30	0.32	0.46
Item 22	0.30	0.35	0.26	0.35	0.29	0.27	0.32	0.49	0.50	0.47	0.63	0.52	0.45	0.49
Item 23	0.27	0.27	0.26	0.21	0.24	0.26	0.29	0.40	0.47	0.34	0.54	0.39	0.27	0.44
Item 26	0.21	0.24	0.18	0.26	0.20	0.17	0.24	0.28	0.27	0.28	0.58	0.26	0.25	0.26
Item 28	0.27	0.33	0.22	0.31	0.23	0.32	0.26	0.41	0.36	0.45	0.44	0.40	0.43	0.41
Item 29	0.57	0.58	0.55	0.64	0.57	0.63	0.54	0.41	0.43	0.39	0.47	0.35	0.39	0.44
Item 30	0.68	0.69	0.66	0.76	0.70	0.69	0.66	0.46	0.45	0.46	0.56	0.34	0.46	0.50
Item 31	0.63	0.66	0.61	0.57	0.61	0.65	0.64	0.31	0.31	0.30	0.31	0.20	0.41	0.32
Item 32	0.31	0.31	0.30	0.36	0.30	0.31	0.31	0.36	0.38	0.33	0.34	0.41	0.39	0.33
Item 33	0.73	0.72	0.73	0.80	0.76	0.76	0.70	0.50	0.48	0.52	0.60	0.43	0.56	0.49
Item 35	0.45	0.43	0.47	0.43	0.44	0.42	0.48	0.40	0.40	0.42	0.49	0.41	0.45	0.38
Item 36	0.39	0.39	0.39	0.41	0.41	0.43	0.38	0.40	0.42	0.38	0.39	0.39	0.44	0.39
Item 37	0.17	0.17	0.17	0.20	0.13	0.12	0.20	0.33	0.30	0.36	0.80	0.42	0.25	0.29
Item 39	0.31	0.32	0.31	0.45	0.26	0.26	0.35	0.56	0.55	0.57	0.54	0.57	0.57	0.55
Item 40	0.32	0.34	0.29	0.32	0.27	0.31	0.33	0.16	0.18	0.13	0.12	0.20	0.14	0.17
Item 44	0.23	0.26	0.20	0.29	0.25	0.23	0.23	0.37	0.39	0.33	0.61	0.36	0.46	0.33
Item 50	0.46	0.50	0.42	0.43	0.51	0.42	0.48	0.31	0.37	0.22	0.40	0.26	0.21	0.35
Item 60	0.54	0.52	0.57	0.57	0.54	0.49	0.57	0.56	0.51	0.64	0.60	0.47	0.47	0.62
Item 61	0.76	0.72	0.80	0.79	0.76	0.75	0.77	0.57	0.56	0.61	0.48	0.47	0.55	0.59
Item 62	0.55	0.60	0.51	0.62	0.50	0.49	0.61	0.49	0.50	0.46	0.41	0.46	0.42	0.49
Item 63	0.56	0.60	0.53	0.59	0.50	0.57	0.57	0.41	0.46	0.35	0.50	0.44	0.46	0.39
Item 64	0.33	0.38	0.28	0.33	0.29	0.32	0.35	0.55	0.51	0.59	0.68	0.54	0.46	0.58
Item 65	0.23	0.24	0.23	0.18	0.24	0.21	0.25	0.28	0.29	0.27	0.03	0.25	0.10	0.37
Item 66	0.68	0.62	0.73	0.70	0.68	0.68	0.67	0.47	0.45	0.53	0.63	0.51	0.37	0.49
Mean	0.44	0.45	0.42	0.47	0.43	0.42	0.45	0.42	0.42	0.41	0.50	0.39	0.39	0.43
S.D.	0.17	0.16	0.17	0.17	0.17	0.18	0.17	0.09	0.08	0.11	0.15	0.09	0.10	0.09
N	3199	1570	1629	105	626	695	1690							
Wtd N	545728	268995	276733	12466	81354	140753	294386							
% Answering Last Item	97%	97%	98%	99%	97%	96%	98%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

BEST COPY AVAILABLE

113

Appendix B-3
Math: First Follow-up (Middle Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.67	0.65	0.70	0.70	0.61	0.54	0.71	0.51	0.53	0.49	0.54	0.56	0.42	0.49
Item 5	0.62	0.61	0.63	0.70	0.58	0.56	0.64	0.56	0.62	0.51	0.61	0.56	0.58	0.55
Item 6	0.75	0.76	0.74	0.78	0.64	0.65	0.79	0.61	0.59	0.63	0.73	0.64	0.65	0.55
Item 8	0.71	0.66	0.75	0.81	0.66	0.68	0.72	0.63	0.63	0.64	0.65	0.70	0.67	0.62
Item 9	0.72	0.72	0.73	0.77	0.66	0.67	0.74	0.59	0.58	0.59	0.47	0.59	0.57	0.58
Item 11	0.70	0.67	0.72	0.79	0.61	0.63	0.72	0.70	0.70	0.69	0.78	0.70	0.67	0.69
Item 12	0.62	0.61	0.63	0.68	0.53	0.60	0.64	0.62	0.65	0.59	0.74	0.55	0.58	0.63
Item 13	0.53	0.53	0.53	0.69	0.47	0.44	0.56	0.53	0.52	0.55	0.39	0.52	0.43	0.54
Item 15	0.49	0.46	0.53	0.52	0.40	0.42	0.52	0.63	0.64	0.62	0.61	0.63	0.61	0.62
Item 16	0.56	0.52	0.60	0.69	0.53	0.50	0.57	0.61	0.63	0.60	0.74	0.64	0.63	0.60
Item 17	0.56	0.57	0.54	0.51	0.51	0.40	0.60	0.52	0.54	0.51	0.49	0.54	0.38	0.53
Item 18	0.47	0.45	0.49	0.47	0.46	0.46	0.48	0.26	0.29	0.22	0.46	0.16	0.24	0.27
Item 22	0.62	0.63	0.62	0.69	0.52	0.55	0.65	0.61	0.63	0.60	0.51	0.65	0.70	0.59
Item 23	0.49	0.49	0.50	0.46	0.36	0.43	0.53	0.54	0.56	0.51	0.53	0.43	0.60	0.53
Item 24	0.49	0.48	0.51	0.60	0.43	0.41	0.51	0.45	0.43	0.48	0.47	0.39	0.38	0.46
Item 25	0.41	0.37	0.44	0.50	0.39	0.38	0.41	0.49	0.46	0.51	0.47	0.49	0.46	0.49
Item 26	0.49	0.50	0.47	0.62	0.45	0.36	0.51	0.60	0.59	0.61	0.57	0.60	0.53	0.59
Item 28	0.58	0.61	0.55	0.68	0.51	0.52	0.60	0.62	0.66	0.59	0.63	0.60	0.65	0.61
Item 30	0.82	0.83	0.82	0.83	0.81	0.82	0.83	0.46	0.52	0.41	0.45	0.57	0.50	0.45
Item 31	0.75	0.77	0.73	0.75	0.69	0.68	0.77	0.39	0.43	0.36	0.46	0.37	0.35	0.39
Item 32	0.59	0.58	0.61	0.65	0.54	0.51	0.62	0.61	0.61	0.60	0.65	0.65	0.50	0.61
Item 33	0.88	0.88	0.89	0.91	0.87	0.88	0.89	0.61	0.63	0.60	0.62	0.65	0.61	0.62
Item 34	0.49	0.52	0.45	0.52	0.48	0.46	0.49	0.23	0.27	0.19	0.37	0.26	0.18	0.22
Item 35	0.69	0.69	0.69	0.71	0.67	0.63	0.70	0.47	0.49	0.45	0.56	0.54	0.51	0.43
Item 36	0.58	0.57	0.60	0.64	0.55	0.56	0.59	0.46	0.51	0.41	0.56	0.44	0.44	0.46
Item 37	0.46	0.45	0.47	0.61	0.37	0.36	0.49	0.65	0.64	0.67	0.76	0.63	0.68	0.63
Item 38	0.59	0.61	0.57	0.50	0.47	0.32	0.66	0.60	0.60	0.59	0.54	0.65	0.60	0.56
Item 39	0.62	0.57	0.66	0.74	0.54	0.55	0.64	0.65	0.65	0.66	0.63	0.68	0.69	0.63
Item 40	0.39	0.44	0.35	0.44	0.37	0.33	0.41	0.30	0.28	0.32	0.42	0.28	0.24	0.28
Item 44	0.49	0.52	0.46	0.59	0.45	0.38	0.51	0.51	0.51	0.52	0.72	0.59	0.43	0.48
Item 50	0.61	0.64	0.58	0.65	0.50	0.55	0.64	0.43	0.46	0.40	0.46	0.42	0.36	0.42
Item 51	0.42	0.46	0.39	0.40	0.31	0.31	0.47	0.49	0.51	0.48	0.62	0.32	0.49	0.50
Item 53	0.55	0.52	0.58	0.60	0.48	0.50	0.57	0.53	0.56	0.49	0.66	0.49	0.52	0.52
Item 54	0.35	0.36	0.33	0.39	0.32	0.25	0.37	0.35	0.31	0.40	0.43	0.27	0.26	0.36
Item 55	0.34	0.32	0.36	0.41	0.29	0.27	0.35	0.40	0.40	0.40	0.58	0.39	0.37	0.40
Item 56	0.29	0.31	0.27	0.29	0.26	0.22	0.31	0.34	0.38	0.29	0.37	0.33	0.16	0.35
Item 57	0.29	0.29	0.30	0.33	0.23	0.22	0.32	0.49	0.49	0.50	0.37	0.50	0.50	0.48
Item 60	0.78	0.76	0.79	0.84	0.73	0.70	0.80	0.66	0.67	0.65	0.68	0.63	0.74	0.63
Item 61	0.91	0.89	0.93	0.92	0.90	0.90	0.92	0.63	0.64	0.61	0.81	0.65	0.61	0.62
Item 63	0.73	0.72	0.73	0.70	0.71	0.65	0.74	0.29	0.32	0.27	0.54	0.36	0.34	0.25
Mean	0.58	0.58	0.58	0.63	0.52	0.51	0.60	0.52	0.53	0.50	0.57	0.52	0.50	0.51
S.D.	0.15	0.15	0.16	0.15	0.16	0.17	0.15	0.12	0.12	0.13	0.12	0.14	0.15	0.12
N	9780	4873	4907	543	1339	1003	6728							
Wtd N	1635418	825367	810051	58936	180723	201679	1166604							
% Answering Last Item	94%	94%	94%	92%	90%	90%	96%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix B-4
Math: First Follow-up (High Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.92	0.92	0.92	0.92	0.79	0.89	0.93	0.56	0.60	0.53	0.64	0.53	0.64	0.55
Item 3	0.93	0.96	0.90	0.94	0.91	0.76	0.94	0.52	0.50	0.51	0.34	0.21	0.69	0.49
Item 5	0.90	0.90	0.89	0.95	0.87	0.94	0.89	0.55	0.51	0.58	0.63	0.62	0.79	0.55
Item 8	0.94	0.94	0.94	0.93	0.93	0.92	0.94	0.66	0.63	0.70	0.41	0.88	0.35	0.68
Item 9	0.95	0.95	0.94	0.96	0.93	0.91	0.95	0.68	0.59	0.77	0.35	0.50	0.92	0.69
Item 11	0.96	0.97	0.95	0.97	0.91	0.80	0.96	0.93	0.85	0.98	0.49	1.11	1.02	0.85
Item 12	0.93	0.92	0.93	0.97	0.86	0.92	0.93	0.75	0.73	0.78	0.60	1.03	0.76	0.73
Item 13	0.87	0.88	0.86	0.94	0.85	0.69	0.88	0.56	0.51	0.59	0.58	0.52	0.84	0.51
Item 15	0.88	0.88	0.88	0.91	0.82	0.84	0.88	0.58	0.56	0.61	0.61	0.86	0.54	0.57
Item 16	0.84	0.85	0.84	0.89	0.78	0.64	0.86	0.54	0.50	0.57	0.63	0.68	0.75	0.46
Item 17	0.84	0.87	0.81	0.90	0.84	0.78	0.84	0.45	0.43	0.46	0.40	0.79	0.47	0.42
Item 18	0.79	0.79	0.78	0.83	0.69	0.86	0.79	0.37	0.42	0.33	0.56	0.35	0.13	0.38
Item 22	0.90	0.92	0.89	0.92	0.87	0.87	0.91	0.60	0.62	0.58	0.74	0.49	0.61	0.60
Item 23	0.87	0.89	0.86	0.86	0.81	0.75	0.88	0.58	0.60	0.56	0.48	0.57	0.51	0.59
Item 24	0.83	0.85	0.82	0.84	0.79	0.82	0.83	0.52	0.56	0.48	0.73	0.86	0.02	0.54
Item 25	0.73	0.73	0.74	0.86	0.75	0.63	0.73	0.53	0.53	0.53	0.70	0.71	0.78	0.49
Item 26	0.84	0.84	0.83	0.88	0.86	0.66	0.84	0.58	0.54	0.62	0.51	0.73	0.88	0.53
Item 28	0.92	0.93	0.92	0.97	0.90	0.90	0.93	0.70	0.80	0.61	0.94	0.99	0.66	0.66
Item 29	0.96	0.97	0.96	0.98	0.97	0.98	0.96	0.73	0.78	0.68	0.82	0.78	0.13	0.75
Item 32	0.93	0.94	0.91	0.96	0.89	0.73	0.94	0.76	0.71	0.79	0.94	0.99	0.89	0.68
Item 34	0.71	0.72	0.71	0.75	0.72	0.75	0.71	0.41	0.40	0.42	0.38	0.30	0.32	0.43
Item 35	0.88	0.89	0.87	0.88	0.87	0.86	0.88	0.41	0.43	0.39	0.53	0.54	0.61	0.38
Item 36	0.85	0.87	0.84	0.88	0.81	0.84	0.85	0.52	0.53	0.50	0.54	0.71	0.56	0.50
Item 37	0.92	0.91	0.92	0.95	0.84	0.88	0.92	0.65	0.69	0.63	0.60	0.89	0.58	0.62
Item 38	0.92	0.94	0.91	0.87	0.92	0.83	0.93	0.56	0.62	0.50	0.78	0.60	0.49	0.52
Item 39	0.92	0.92	0.93	0.97	0.89	0.80	0.93	0.62	0.58	0.68	0.66	0.71	0.74	0.57
Item 40	0.66	0.71	0.61	0.82	0.62	0.46	0.66	0.55	0.57	0.53	0.53	0.59	0.43	0.55
Item 43	0.31	0.34	0.28	0.34	0.24	0.24	0.32	0.38	0.37	0.38	0.50	0.40	0.34	0.37
Item 44	0.86	0.87	0.84	0.91	0.77	0.79	0.86	0.55	0.56	0.54	0.69	0.55	0.43	0.55
Item 46	0.55	0.59	0.51	0.65	0.46	0.42	0.56	0.52	0.56	0.48	0.60	0.41	0.40	0.52
Item 47	0.45	0.46	0.45	0.51	0.38	0.38	0.46	0.35	0.34	0.36	0.35	0.23	0.49	0.34
Item 49	0.66	0.64	0.68	0.80	0.54	0.58	0.67	0.59	0.61	0.58	0.75	0.52	0.73	0.57
Item 50	0.86	0.90	0.83	0.86	0.86	0.71	0.87	0.49	0.43	0.52	0.39	0.68	0.88	0.43
Item 51	0.77	0.79	0.74	0.80	0.73	0.51	0.78	0.55	0.55	0.55	0.76	0.63	0.76	0.50
Item 52	0.53	0.53	0.53	0.68	0.47	0.44	0.52	0.62	0.63	0.61	0.60	0.60	0.64	0.62
Item 53	0.83	0.83	0.83	0.91	0.82	0.81	0.83	0.51	0.52	0.49	0.61	0.13	0.42	0.52
Item 54	0.69	0.72	0.67	0.79	0.73	0.56	0.70	0.67	0.67	0.66	0.78	0.69	0.75	0.65
Item 55	0.68	0.70	0.66	0.74	0.60	0.47	0.69	0.56	0.56	0.57	0.68	0.50	0.74	0.54
Item 56	0.60	0.65	0.54	0.68	0.56	0.45	0.60	0.48	0.46	0.49	0.53	0.43	0.37	0.48
Item 57	0.64	0.68	0.61	0.71	0.56	0.51	0.65	0.53	0.51	0.54	0.48	0.54	0.63	0.51
Mean	0.80	0.81	0.79	0.85	0.76	0.71	0.80	0.57	0.56	0.57	0.60	0.62	0.59	0.55
S.D.	0.15	0.15	0.16	0.13	0.17	0.18	0.15	0.11	0.11	0.12	0.15	0.22	0.23	0.11
N	4814	2452	2362	463	234	154	3941							
Wtd N	689739	348482	341258	33657	27560	28924	595902							
% Answering Last Item	97%	97%	98%	98%	94%	96%	97%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix B-5
Math: Second Follow-up (Low Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.52	0.50	0.54	0.46	0.49	0.40	0.57	0.42	0.48	0.38	0.35	0.24	0.43	0.45
Item 2	0.46	0.48	0.43	0.43	0.46	0.42	0.47	0.45	0.45	0.43	0.11	0.56	0.53	0.41
Item 3	0.58	0.62	0.54	0.54	0.47	0.52	0.63	0.40	0.40	0.38	0.49	0.40	0.44	0.36
Item 4	0.90	0.90	0.89	0.88	0.90	0.89	0.90	0.53	0.61	0.45	0.45	0.32	0.59	0.56
Item 6	0.58	0.64	0.51	0.59	0.54	0.50	0.64	0.48	0.49	0.44	0.43	0.50	0.36	0.50
Item 7	0.57	0.66	0.47	0.58	0.54	0.30	0.69	0.48	0.46	0.48	0.60	0.43	0.42	0.47
Item 8	0.44	0.45	0.43	0.44	0.54	0.41	0.44	0.43	0.42	0.44	0.60	0.51	0.42	0.41
Item 9	0.48	0.54	0.42	0.52	0.44	0.45	0.52	0.47	0.52	0.39	0.49	0.42	0.44	0.48
Item 11	0.42	0.44	0.38	0.40	0.42	0.46	0.41	0.50	0.45	0.55	0.70	0.47	0.61	0.48
Item 12	0.40	0.40	0.41	0.34	0.37	0.39	0.43	0.50	0.54	0.45	0.49	0.39	0.50	0.50
Item 13	0.35	0.36	0.33	0.40	0.34	0.33	0.36	0.31	0.35	0.26	0.53	0.29	0.22	0.33
Item 14	0.80	0.80	0.79	0.84	0.84	0.80	0.78	0.46	0.53	0.37	0.44	0.42	0.50	0.48
Item 21	0.51	0.49	0.54	0.56	0.56	0.56	0.47	0.43	0.43	0.45	0.46	0.40	0.54	0.45
Item 22	0.31	0.34	0.27	0.36	0.32	0.23	0.33	0.44	0.45	0.41	0.69	0.51	0.38	0.42
Item 23	0.37	0.36	0.38	0.32	0.29	0.39	0.38	0.38	0.40	0.37	0.19	0.39	0.38	0.40
Item 26	0.22	0.22	0.21	0.31	0.21	0.22	0.22	0.32	0.29	0.36	0.42	0.45	0.27	0.31
Item 28	0.31	0.37	0.24	0.32	0.26	0.28	0.33	0.50	0.47	0.51	0.57	0.48	0.45	0.50
Item 29	0.56	0.59	0.53	0.60	0.56	0.54	0.57	0.37	0.36	0.37	0.08	0.31	0.26	0.44
Item 30	0.75	0.74	0.75	0.76	0.80	0.73	0.74	0.23	0.23	0.23	0.57	0.27	0.11	0.26
Item 31	0.66	0.67	0.64	0.72	0.59	0.66	0.67	0.33	0.38	0.28	0.39	0.24	0.36	0.36
Item 32	0.35	0.38	0.32	0.42	0.32	0.33	0.35	0.44	0.52	0.33	0.54	0.46	0.47	0.46
Item 33	0.74	0.72	0.76	0.79	0.81	0.80	0.70	0.35	0.32	0.40	0.40	0.38	0.33	0.39
Item 34	0.43	0.44	0.42	0.33	0.47	0.43	0.42	0.21	0.18	0.24	0.32	0.11	0.32	0.21
Item 35	0.43	0.42	0.45	0.43	0.43	0.34	0.48	0.34	0.31	0.39	0.45	0.32	0.39	0.30
Item 36	0.41	0.40	0.42	0.33	0.48	0.35	0.43	0.37	0.40	0.35	0.48	0.33	0.36	0.37
Item 37	0.20	0.22	0.18	0.14	0.16	0.21	0.20	0.36	0.33	0.39	0.67	0.45	0.16	0.42
Item 39	0.34	0.34	0.34	0.42	0.40	0.35	0.32	0.55	0.51	0.61	0.73	0.51	0.59	0.56
Item 44	0.26	0.29	0.24	0.22	0.27	0.17	0.31	0.41	0.40	0.41	0.71	0.29	0.42	0.40
Item 45	0.25	0.25	0.25	0.22	0.22	0.22	0.27	0.16	0.09	0.25	0.09	0.14	0.29	0.10
Item 50	0.44	0.50	0.38	0.48	0.45	0.39	0.47	0.35	0.45	0.19	0.05	0.30	0.32	0.37
Item 60	0.65	0.65	0.65	0.70	0.68	0.59	0.67	0.65	0.66	0.65	0.38	0.64	0.65	0.67
Item 61	0.85	0.83	0.86	0.77	0.85	0.86	0.84	0.58	0.62	0.54	0.76	0.41	0.63	0.61
Item 62	0.66	0.70	0.62	0.60	0.64	0.56	0.72	0.50	0.48	0.51	0.57	0.54	0.43	0.50
Item 63	0.59	0.60	0.58	0.48	0.54	0.56	0.61	0.44	0.42	0.45	0.57	0.37	0.43	0.44
Item 64	0.32	0.37	0.27	0.31	0.29	0.26	0.36	0.50	0.43	0.57	0.67	0.56	0.32	0.52
Item 66	0.80	0.78	0.81	0.82	0.81	0.79	0.80	0.45	0.49	0.43	0.48	0.44	0.36	0.50
Item 67	0.60	0.58	0.63	0.56	0.55	0.56	0.65	0.43	0.45	0.43	0.44	0.22	0.39	0.49
Item 68	0.14	0.15	0.14	0.14	0.13	0.11	0.16	0.37	0.45	0.27	0.20	0.25	0.24	0.43
Item 69	0.28	0.31	0.23	0.34	0.27	0.17	0.33	0.38	0.40	0.32	0.38	0.33	0.22	0.39
Item 70	0.22	0.25	0.19	0.23	0.20	0.23	0.22	0.28	0.29	0.25	0.60	0.35	0.15	0.33
Mean	0.48	0.49	0.46	0.48	0.47	0.44	0.50	0.41	0.42	0.40	0.46	0.39	0.39	0.43
S.D.	0.19	0.19	0.20	0.19	0.20	0.20	0.19	0.10	0.11	0.10	0.19	0.11	0.13	0.10
N	2554	1293	1261	93	473	533	1395							
Wtd N	429799	224020	205779	9790	60546	99993	245208							
% Answering Last Item	98%	97%	98%	94%	96%	97%	99%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix B-6
Math: Second Follow-up (Middle Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.76	0.74	0.77	0.77	0.70	0.66	0.78	0.54	0.53	0.56	0.60	0.51	0.58	0.52
Item 9	0.78	0.79	0.77	0.84	0.71	0.69	0.81	0.61	0.60	0.63	0.74	0.61	0.52	0.61
Item 11	0.78	0.79	0.77	0.87	0.69	0.70	0.81	0.72	0.71	0.72	0.77	0.73	0.75	0.69
Item 12	0.74	0.74	0.74	0.79	0.66	0.66	0.77	0.65	0.64	0.66	0.72	0.58	0.66	0.65
Item 22	0.73	0.74	0.72	0.82	0.66	0.67	0.75	0.55	0.57	0.53	0.55	0.59	0.59	0.51
Item 23	0.60	0.59	0.61	0.56	0.48	0.48	0.64	0.60	0.63	0.56	0.66	0.50	0.60	0.58
Item 24	0.53	0.55	0.51	0.55	0.46	0.43	0.56	0.54	0.55	0.53	0.54	0.53	0.45	0.55
Item 25	0.46	0.42	0.51	0.62	0.45	0.40	0.47	0.49	0.50	0.49	0.42	0.47	0.57	0.48
Item 26	0.56	0.58	0.54	0.64	0.53	0.42	0.59	0.57	0.57	0.58	0.57	0.62	0.47	0.57
Item 28	0.66	0.69	0.63	0.74	0.55	0.61	0.68	0.63	0.64	0.62	0.72	0.69	0.60	0.61
Item 30	0.86	0.87	0.86	0.89	0.86	0.84	0.87	0.36	0.38	0.33	0.41	0.53	0.37	0.31
Item 31	0.77	0.79	0.75	0.79	0.71	0.71	0.79	0.43	0.49	0.37	0.45	0.45	0.46	0.40
Item 32	0.69	0.69	0.68	0.78	0.61	0.58	0.72	0.62	0.66	0.58	0.67	0.66	0.67	0.58
Item 33	0.90	0.89	0.90	0.95	0.86	0.86	0.91	0.44	0.44	0.44	0.45	0.43	0.61	0.38
Item 34	0.58	0.58	0.57	0.62	0.56	0.49	0.59	0.37	0.36	0.38	0.52	0.32	0.35	0.36
Item 35	0.75	0.74	0.75	0.74	0.69	0.69	0.77	0.45	0.49	0.42	0.44	0.56	0.43	0.42
Item 36	0.64	0.64	0.64	0.64	0.61	0.60	0.66	0.48	0.51	0.45	0.48	0.48	0.52	0.47
Item 37	0.50	0.49	0.52	0.65	0.39	0.39	0.54	0.64	0.66	0.62	0.62	0.66	0.53	0.64
Item 39	0.72	0.69	0.75	0.85	0.65	0.63	0.74	0.71	0.71	0.72	0.65	0.65	0.75	0.70
Item 40	0.39	0.45	0.33	0.40	0.36	0.30	0.41	0.37	0.41	0.31	0.51	0.29	0.17	0.39
Item 41	0.27	0.29	0.26	0.31	0.25	0.26	0.28	0.20	0.23	0.15	0.50	0.15	0.07	0.21
Item 43	0.20	0.21	0.18	0.20	0.14	0.12	0.22	0.33	0.28	0.37	0.50	0.29	0.01	0.33
Item 44	0.58	0.61	0.55	0.72	0.50	0.46	0.60	0.61	0.61	0.60	0.54	0.62	0.59	0.59
Item 45	0.31	0.29	0.33	0.29	0.28	0.24	0.33	0.34	0.35	0.34	0.35	0.35	0.37	0.32
Item 50	0.67	0.70	0.63	0.72	0.57	0.54	0.71	0.46	0.48	0.45	0.53	0.46	0.43	0.43
Item 51	0.56	0.59	0.54	0.55	0.45	0.42	0.61	0.61	0.64	0.58	0.66	0.62	0.58	0.59
Item 54	0.36	0.40	0.33	0.40	0.29	0.26	0.39	0.49	0.51	0.46	0.63	0.55	0.40	0.46
Item 55	0.36	0.34	0.38	0.43	0.30	0.30	0.38	0.45	0.49	0.41	0.52	0.41	0.38	0.46
Item 56	0.33	0.35	0.30	0.43	0.26	0.24	0.35	0.42	0.43	0.40	0.47	0.42	0.38	0.39
Item 57	0.36	0.37	0.36	0.43	0.30	0.29	0.39	0.53	0.54	0.52	0.56	0.61	0.26	0.55
Item 58	0.06	0.06	0.05	0.09	0.02	0.05	0.06	0.25	0.26	0.23	0.63	0.21	-0.11	0.27
Item 59	0.15	0.16	0.15	0.18	0.12	0.11	0.17	0.17	0.19	0.14	0.33	0.17	-0.01	0.16
Item 60	0.91	0.90	0.92	0.94	0.87	0.85	0.93	0.79	0.77	0.82	0.75	0.67	0.91	0.76
Item 61	0.93	0.91	0.94	0.94	0.90	0.86	0.94	0.59	0.58	0.62	0.75	0.58	0.69	0.51
Item 63	0.73	0.73	0.74	0.76	0.70	0.70	0.74	0.30	0.33	0.27	0.33	0.48	0.30	0.25
Item 69	0.40	0.45	0.34	0.44	0.38	0.30	0.42	0.39	0.42	0.35	0.51	0.36	0.29	0.39
Item 70	0.45	0.46	0.43	0.50	0.37	0.33	0.48	0.60	0.60	0.60	0.76	0.59	0.58	0.58
Item 71	0.46	0.49	0.43	0.49	0.50	0.41	0.46	0.22	0.23	0.21	0.19	0.23	0.26	0.22
Item 72	0.33	0.33	0.34	0.34	0.33	0.31	0.34	0.25	0.23	0.26	0.22	0.20	0.16	0.27
Item 73	0.23	0.23	0.22	0.30	0.20	0.16	0.24	0.52	0.51	0.52	0.55	0.54	0.36	0.53
Mean	0.55	0.56	0.54	0.60	0.50	0.48	0.57	0.48	0.49	0.47	0.54	0.49	0.44	0.47
S.D.	0.22	0.22	0.23	0.23	0.22	0.22	0.22	0.15	0.15	0.16	0.14	0.15	0.22	0.15
N	7717	3746	3971	482	1087	758	5269							
Wtd N	1293720	652015	641705	53853	151143	169234	901264							
% Answering Last Item	91%	92%	90%	91%	87%	87%	92%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

BEST COPY AVAILABLE

Appendix B-7
Math: Second Follow-up (High Form)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 24	0.90	0.91	0.89	0.87	0.92	0.87	0.90	0.50	0.53	0.48	0.56	0.52	0.24	0.52
Item 25	0.82	0.82	0.82	0.92	0.83	0.83	0.81	0.40	0.49	0.30	0.45	0.46	0.27	0.40
Item 26	0.86	0.87	0.85	0.91	0.91	0.74	0.86	0.37	0.41	0.32	0.38	0.13	0.28	0.38
Item 27	0.40	0.43	0.37	0.58	0.33	0.30	0.40	0.55	0.55	0.54	0.69	0.55	0.58	0.52
Item 36	0.89	0.89	0.90	0.91	0.86	0.89	0.90	0.46	0.46	0.48	0.45	0.07	0.48	0.49
Item 37	0.95	0.94	0.96	0.96	0.94	0.91	0.95	0.43	0.50	0.34	0.42	0.51	0.16	0.44
Item 39	0.97	0.96	0.98	0.98	0.99	0.95	0.97	0.41	0.50	0.29	0.62	0.46	0.00	0.43
Item 40	0.80	0.85	0.74	0.89	0.80	0.72	0.80	0.63	0.60	0.64	0.64	0.63	0.38	0.63
Item 41	0.48	0.51	0.44	0.56	0.39	0.38	0.48	0.49	0.50	0.47	0.57	0.59	0.31	0.49
Item 42	0.51	0.54	0.47	0.60	0.42	0.38	0.51	0.48	0.50	0.45	0.41	0.67	0.54	0.47
Item 43	0.41	0.43	0.39	0.51	0.29	0.30	0.42	0.40	0.38	0.42	0.40	0.80	0.26	0.38
Item 44	0.92	0.93	0.92	0.95	0.89	0.93	0.92	0.51	0.56	0.45	0.34	0.60	0.55	0.51
Item 45	0.53	0.54	0.52	0.61	0.45	0.44	0.53	0.38	0.44	0.30	0.50	0.50	0.37	0.36
Item 46	0.71	0.74	0.67	0.77	0.67	0.55	0.71	0.55	0.62	0.47	0.63	0.66	0.66	0.53
Item 47	0.59	0.61	0.56	0.63	0.57	0.55	0.59	0.37	0.41	0.33	0.36	0.47	0.45	0.36
Item 48	0.46	0.49	0.42	0.60	0.39	0.32	0.46	0.58	0.56	0.61	0.68	0.60	0.43	0.58
Item 49	0.90	0.90	0.91	0.95	0.92	0.92	0.90	0.68	0.66	0.72	0.70	0.54	0.76	0.69
Item 51	0.91	0.92	0.90	0.89	0.93	0.79	0.92	0.58	0.60	0.55	0.65	0.39	0.55	0.59
Item 52	0.76	0.74	0.77	0.87	0.71	0.76	0.75	0.65	0.64	0.69	0.73	0.36	0.66	0.67
Item 54	0.81	0.83	0.79	0.83	0.82	0.78	0.81	0.57	0.56	0.57	0.82	0.52	0.33	0.57
Item 55	0.76	0.77	0.75	0.83	0.70	0.73	0.77	0.58	0.58	0.59	0.82	0.54	0.71	0.56
Item 56	0.71	0.75	0.67	0.75	0.75	0.62	0.71	0.44	0.42	0.45	0.31	0.48	0.76	0.44
Item 57	0.79	0.80	0.79	0.83	0.74	0.71	0.80	0.51	0.45	0.58	0.85	0.36	0.26	0.50
Item 58	0.15	0.17	0.12	0.24	0.15	0.09	0.14	0.56	0.60	0.47	0.45	0.53	0.42	0.57
Item 59	0.24	0.27	0.21	0.33	0.16	0.14	0.25	0.48	0.51	0.42	0.61	0.56	0.11	0.47
Item 67	0.93	0.92	0.95	0.93	0.93	0.89	0.94	0.44	0.48	0.42	0.89	0.52	0.42	0.41
Item 68	0.89	0.90	0.88	0.89	0.88	0.79	0.90	0.61	0.60	0.61	0.71	0.57	0.68	0.60
Item 69	0.67	0.71	0.63	0.68	0.62	0.47	0.68	0.45	0.42	0.46	0.44	0.25	0.54	0.46
Item 70	0.84	0.87	0.81	0.88	0.88	0.79	0.84	0.51	0.45	0.57	0.58	0.51	0.07	0.52
Item 71	0.59	0.65	0.54	0.64	0.57	0.52	0.60	0.35	0.37	0.31	0.25	0.40	0.43	0.35
Item 72	0.57	0.59	0.55	0.69	0.56	0.44	0.57	0.48	0.48	0.47	0.55	0.41	0.54	0.47
Item 73	0.57	0.58	0.56	0.66	0.56	0.49	0.57	0.59	0.67	0.49	0.75	0.56	0.58	0.58
Item 74	0.41	0.44	0.38	0.50	0.47	0.29	0.41	0.40	0.37	0.44	0.46	0.27	0.31	0.41
Item 75	0.54	0.51	0.58	0.66	0.58	0.50	0.53	0.54	0.57	0.53	0.61	0.50	0.42	0.54
Item 76	0.41	0.40	0.43	0.61	0.31	0.33	0.41	0.65	0.68	0.63	0.72	0.58	0.53	0.65
Item 77	0.37	0.39	0.34	0.60	0.26	0.35	0.36	0.61	0.63	0.57	0.70	0.32	0.40	0.62
Item 78	0.16	0.17	0.15	0.21	0.14	0.07	0.16	0.43	0.48	0.36	0.51	0.38	0.06	0.43
Item 79	0.30	0.31	0.28	0.38	0.29	0.28	0.29	0.44	0.41	0.47	0.51	0.61	0.24	0.43
Item 80	0.23	0.26	0.20	0.33	0.19	0.16	0.23	0.64	0.64	0.61	0.42	0.69	0.63	0.64
Item 81	0.26	0.33	0.18	0.40	0.24	0.19	0.26	0.59	0.60	0.53	0.58	0.53	0.64	0.58
Mean	0.62	0.64	0.61	0.70	0.60	0.55	0.63	0.51	0.52	0.49	0.57	0.49	0.42	0.51
S.D.	0.24	0.24	0.25	0.21	0.27	0.26	0.25	0.09	0.09	0.11	0.16	0.14	0.20	0.09
N	3965	2087	1878	370	200	115	3264							
Wtd N	557388	293382	264007	27522	24771	20684	482351							
% Answering Last Item	81%	82%	79%	87%	69%	67%	82%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix C: Science

Appendix C-1
Science: Base Year (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.70	0.69	0.70	0.69	0.63	0.51	0.75	0.57	0.60	0.55	0.57	0.48	0.45	0.57
Item 2	0.79	0.80	0.77	0.81	0.72	0.69	0.81	0.51	0.61	0.42	0.55	0.50	0.45	0.50
Item 3	0.64	0.65	0.64	0.68	0.57	0.53	0.67	0.48	0.49	0.47	0.52	0.46	0.39	0.47
Item 4	0.67	0.63	0.70	0.66	0.62	0.57	0.69	0.45	0.47	0.45	0.42	0.37	0.40	0.45
Item 5	0.76	0.77	0.74	0.77	0.67	0.58	0.80	0.71	0.78	0.65	0.71	0.64	0.63	0.72
Item 6	0.76	0.76	0.76	0.76	0.65	0.65	0.80	0.67	0.72	0.63	0.70	0.61	0.59	0.68
Item 7	0.65	0.70	0.61	0.70	0.61	0.55	0.68	0.50	0.57	0.42	0.45	0.48	0.45	0.49
Item 8	0.57	0.61	0.54	0.53	0.48	0.48	0.61	0.46	0.49	0.42	0.52	0.45	0.38	0.45
Item 9	0.64	0.64	0.65	0.67	0.56	0.53	0.68	0.51	0.52	0.51	0.53	0.47	0.46	0.50
Item 10	0.53	0.54	0.53	0.55	0.41	0.43	0.57	0.53	0.55	0.51	0.58	0.45	0.38	0.53
Item 11	0.48	0.50	0.46	0.53	0.42	0.40	0.50	0.41	0.46	0.36	0.40	0.44	0.36	0.41
Item 12	0.66	0.70	0.63	0.71	0.57	0.52	0.70	0.57	0.60	0.55	0.60	0.54	0.48	0.56
Item 13	0.72	0.70	0.75	0.76	0.66	0.61	0.75	0.54	0.59	0.50	0.50	0.52	0.50	0.53
Item 14	0.53	0.58	0.49	0.55	0.36	0.25	0.61	0.65	0.66	0.65	0.68	0.53	0.49	0.64
Item 15	0.39	0.37	0.41	0.45	0.37	0.28	0.41	0.47	0.46	0.48	0.46	0.45	0.42	0.46
Item 16	0.46	0.46	0.46	0.50	0.43	0.39	0.48	0.42	0.43	0.40	0.47	0.31	0.32	0.44
Item 17	0.42	0.45	0.39	0.45	0.34	0.31	0.45	0.49	0.53	0.45	0.54	0.39	0.31	0.51
Item 18	0.45	0.49	0.41	0.45	0.34	0.30	0.50	0.54	0.56	0.51	0.55	0.41	0.34	0.55
Item 19	0.42	0.43	0.41	0.49	0.33	0.31	0.46	0.50	0.52	0.49	0.52	0.39	0.45	0.50
Item 20	0.41	0.41	0.41	0.45	0.36	0.36	0.43	0.35	0.37	0.33	0.44	0.28	0.30	0.36
Item 21	0.42	0.44	0.40	0.47	0.36	0.36	0.44	0.39	0.42	0.34	0.41	0.30	0.27	0.41
Item 22	0.37	0.35	0.39	0.44	0.33	0.29	0.39	0.38	0.40	0.37	0.38	0.31	0.33	0.38
Item 23	0.39	0.40	0.39	0.43	0.35	0.34	0.41	0.27	0.30	0.24	0.35	0.19	0.24	0.27
Item 24	0.33	0.33	0.32	0.34	0.24	0.20	0.36	0.56	0.56	0.56	0.58	0.53	0.50	0.54
Item 25	0.22	0.21	0.23	0.24	0.18	0.16	0.24	0.37	0.35	0.39	0.35	0.33	0.33	0.36
Mean	0.54	0.55	0.53	0.56	0.46	0.42	0.57	0.49	0.52	0.47	0.51	0.43	0.41	0.49
S.D.	0.15	0.16	0.15	0.15	0.15	0.14	0.16	0.10	0.11	0.10	0.10	0.10	0.09	0.10
N	23616	11750	11866	1500	2997	2865	15852							
Wtd N	2889974	1447373	1442602	102242	291843	371291	2072010							
% Answering Last Item	97%	97%	98%	97%	96%	94%	98%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix C-2
Science: First Follow-up (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 3	0.72	0.73	0.71	0.77	0.65	0.58	0.76	0.53	0.55	0.51	0.52	0.38	0.44	0.53
Item 4	0.74	0.72	0.77	0.74	0.71	0.63	0.77	0.51	0.54	0.50	0.51	0.46	0.43	0.51
Item 5	0.78	0.81	0.75	0.78	0.69	0.61	0.82	0.71	0.79	0.64	0.75	0.65	0.62	0.72
Item 6	0.84	0.83	0.84	0.88	0.75	0.73	0.87	0.70	0.76	0.65	0.64	0.58	0.66	0.72
Item 10	0.59	0.60	0.58	0.67	0.48	0.46	0.63	0.60	0.63	0.57	0.60	0.52	0.55	0.60
Item 12	0.73	0.77	0.69	0.80	0.65	0.56	0.77	0.61	0.66	0.55	0.53	0.57	0.50	0.60
Item 14	0.65	0.71	0.59	0.66	0.46	0.37	0.73	0.71	0.73	0.69	0.80	0.61	0.61	0.69
Item 15	0.54	0.51	0.57	0.59	0.54	0.41	0.57	0.49	0.53	0.48	0.46	0.40	0.49	0.49
Item 16	0.56	0.57	0.55	0.59	0.50	0.46	0.58	0.52	0.54	0.50	0.61	0.38	0.43	0.54
Item 17	0.57	0.60	0.53	0.59	0.45	0.38	0.62	0.66	0.69	0.62	0.64	0.57	0.52	0.66
Item 18	0.58	0.62	0.53	0.53	0.41	0.39	0.64	0.61	0.64	0.57	0.62	0.48	0.52	0.59
Item 19	0.54	0.55	0.54	0.61	0.41	0.38	0.59	0.60	0.60	0.61	0.60	0.54	0.53	0.58
Item 20	0.50	0.50	0.49	0.57	0.39	0.41	0.52	0.47	0.47	0.48	0.49	0.48	0.34	0.48
Item 21	0.51	0.54	0.48	0.54	0.44	0.43	0.54	0.49	0.52	0.44	0.53	0.39	0.31	0.52
Item 22	0.46	0.44	0.47	0.54	0.41	0.35	0.48	0.46	0.51	0.42	0.41	0.41	0.42	0.47
Item 23	0.50	0.51	0.49	0.56	0.44	0.45	0.52	0.38	0.39	0.36	0.48	0.30	0.35	0.39
Item 24	0.42	0.45	0.39	0.43	0.33	0.28	0.46	0.59	0.60	0.57	0.58	0.60	0.60	0.56
Item 25	0.32	0.31	0.33	0.36	0.25	0.25	0.34	0.51	0.52	0.50	0.50	0.40	0.45	0.52
Item 26	0.52	0.63	0.42	0.55	0.40	0.29	0.59	0.60	0.62	0.58	0.63	0.53	0.52	0.57
Item 27	0.28	0.30	0.26	0.33	0.21	0.20	0.30	0.55	0.60	0.49	0.76	0.47	0.24	0.58
Item 29	0.49	0.57	0.40	0.51	0.33	0.25	0.55	0.63	0.65	0.60	0.57	0.48	0.45	0.62
Item 30	0.50	0.57	0.42	0.55	0.41	0.28	0.55	0.55	0.59	0.49	0.56	0.53	0.50	0.52
Item 32	0.26	0.30	0.21	0.29	0.17	0.15	0.29	0.56	0.59	0.49	0.68	0.39	0.34	0.56
Item 33	0.56	0.56	0.57	0.61	0.44	0.42	0.60	0.62	0.65	0.59	0.65	0.48	0.46	0.63
Item 34	0.47	0.46	0.48	0.49	0.39	0.40	0.50	0.44	0.45	0.43	0.45	0.44	0.48	0.41
Mean	0.55	0.57	0.52	0.58	0.45	0.40	0.58	0.56	0.59	0.53	0.58	0.48	0.47	0.56
S.D.	0.14	0.14	0.15	0.14	0.15	0.14	0.15	0.08	0.09	0.08	0.10	0.09	0.10	0.08
N	17684	8841	8843	1103	2160	1832	12316							
Wtd N	2849102	1433449	1415653	104278	285180	360731	2050740							
% Answering Last Item	98%	98%	98%	96%	95%	96%	99%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix C-3
Science: Second Follow-up (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 4	0.78	0.75	0.81	0.80	0.74	0.59	0.82	0.53	0.57	0.53	0.46	0.43	0.44	0.53
Item 5	0.81	0.84	0.78	0.82	0.72	0.66	0.86	0.70	0.78	0.63	0.67	0.62	0.59	0.71
Item 6	0.88	0.88	0.87	0.83	0.81	0.78	0.91	0.67	0.71	0.64	0.58	0.65	0.60	0.67
Item 10	0.65	0.66	0.64	0.69	0.52	0.53	0.69	0.65	0.68	0.63	0.76	0.58	0.56	0.65
Item 12	0.73	0.78	0.67	0.76	0.63	0.58	0.77	0.63	0.66	0.59	0.67	0.56	0.53	0.63
Item 14	0.70	0.76	0.64	0.74	0.51	0.43	0.78	0.73	0.73	0.73	0.76	0.68	0.63	0.70
Item 15	0.56	0.53	0.58	0.61	0.56	0.46	0.58	0.47	0.49	0.47	0.48	0.40	0.50	0.47
Item 16	0.58	0.58	0.57	0.65	0.50	0.47	0.61	0.54	0.57	0.52	0.61	0.44	0.44	0.56
Item 17	0.63	0.67	0.60	0.70	0.50	0.43	0.69	0.71	0.73	0.68	0.73	0.62	0.57	0.70
Item 18	0.65	0.69	0.60	0.65	0.53	0.41	0.71	0.61	0.64	0.58	0.50	0.49	0.53	0.59
Item 19	0.59	0.59	0.58	0.61	0.48	0.43	0.63	0.62	0.63	0.63	0.57	0.58	0.49	0.63
Item 22	0.47	0.46	0.48	0.51	0.44	0.31	0.50	0.46	0.49	0.44	0.51	0.36	0.44	0.45
Item 24	0.45	0.46	0.43	0.50	0.34	0.25	0.50	0.62	0.64	0.60	0.54	0.62	0.55	0.59
Item 26	0.61	0.72	0.49	0.59	0.47	0.36	0.68	0.64	0.65	0.64	0.68	0.57	0.57	0.61
Item 27	0.32	0.35	0.29	0.38	0.25	0.19	0.35	0.65	0.68	0.60	0.73	0.57	0.50	0.66
Item 28	0.73	0.75	0.71	0.78	0.68	0.66	0.75	0.52	0.52	0.52	0.65	0.47	0.49	0.53
Item 29	0.58	0.67	0.49	0.60	0.44	0.30	0.65	0.69	0.72	0.65	0.60	0.66	0.50	0.67
Item 30	0.58	0.64	0.52	0.64	0.49	0.34	0.64	0.60	0.63	0.56	0.62	0.57	0.51	0.57
Item 31	0.59	0.58	0.60	0.57	0.51	0.48	0.62	0.50	0.51	0.49	0.42	0.47	0.55	0.47
Item 32	0.34	0.38	0.30	0.38	0.23	0.18	0.39	0.67	0.70	0.62	0.73	0.50	0.43	0.69
Item 33	0.64	0.64	0.65	0.67	0.55	0.47	0.69	0.65	0.69	0.61	0.61	0.57	0.63	0.64
Item 35	0.43	0.43	0.42	0.53	0.38	0.30	0.45	0.56	0.59	0.53	0.53	0.47	0.42	0.58
Item 36	0.43	0.45	0.41	0.42	0.38	0.38	0.45	0.33	0.36	0.30	0.36	0.33	0.25	0.34
Item 37	0.29	0.30	0.28	0.36	0.28	0.25	0.30	0.31	0.31	0.31	0.31	0.30	0.23	0.32
Item 38	0.13	0.13	0.13	0.16	0.10	0.13	0.14	0.26	0.40	0.12	0.24	0.11	0.10	0.33
Mean	0.57	0.59	0.54	0.60	0.48	0.42	0.61	0.57	0.60	0.54	0.57	0.50	0.48	0.57
S.D.	0.17	0.18	0.17	0.16	0.16	0.16	0.18	0.12	0.12	0.13	0.14	0.13	0.12	0.11
N	14134	7070	7064	937	1744	1389	9870							
Wtd N	2262896	1159087	1103809	90180	233539	287625	1617361							
% Answering Last Item	97%	97%	97%	98%	95%	95%	98%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix D: History/Citizenship/Geography

Appendix D-1
History/Citizenship/Geography: Base Year (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.69	0.70	0.67	0.64	0.50	0.54	0.74	0.63	0.67	0.59	0.62	0.56	0.58	0.62
Item 2	0.49	0.51	0.46	0.49	0.37	0.32	0.53	0.53	0.55	0.51	0.48	0.43	0.33	0.54
Item 4	0.48	0.48	0.48	0.57	0.44	0.33	0.51	0.57	0.59	0.55	0.57	0.47	0.44	0.59
Item 5	0.55	0.52	0.58	0.64	0.48	0.46	0.57	0.53	0.54	0.52	0.53	0.50	0.47	0.53
Item 6	0.43	0.47	0.39	0.45	0.34	0.31	0.46	0.48	0.51	0.45	0.54	0.35	0.33	0.49
Item 7	0.77	0.77	0.78	0.75	0.64	0.73	0.81	0.66	0.69	0.63	0.73	0.61	0.60	0.67
Item 8	0.58	0.61	0.55	0.59	0.53	0.50	0.61	0.59	0.64	0.54	0.65	0.53	0.49	0.61
Item 9	0.42	0.44	0.40	0.57	0.49	0.35	0.41	0.42	0.43	0.41	0.52	0.43	0.34	0.44
Item 10	0.47	0.48	0.46	0.53	0.40	0.33	0.51	0.60	0.63	0.57	0.60	0.52	0.51	0.61
Item 11	0.45	0.46	0.44	0.54	0.38	0.36	0.47	0.47	0.52	0.42	0.51	0.44	0.33	0.48
Item 13	0.48	0.48	0.47	0.52	0.44	0.40	0.50	0.50	0.53	0.46	0.52	0.46	0.41	0.51
Item 14	0.78	0.78	0.78	0.80	0.70	0.68	0.81	0.59	0.62	0.57	0.63	0.54	0.52	0.59
Item 15	0.66	0.66	0.65	0.65	0.53	0.51	0.70	0.61	0.62	0.59	0.65	0.54	0.50	0.60
Item 16	0.90	0.89	0.91	0.90	0.84	0.82	0.92	0.76	0.79	0.73	0.79	0.67	0.67	0.78
Item 17	0.80	0.79	0.80	0.84	0.74	0.66	0.83	0.58	0.59	0.58	0.59	0.54	0.48	0.58
Item 18	0.24	0.26	0.22	0.28	0.22	0.21	0.25	0.29	0.28	0.31	0.37	0.19	0.07	0.34
Item 19	0.84	0.84	0.83	0.81	0.73	0.83	0.86	0.64	0.67	0.60	0.70	0.65	0.63	0.64
Item 21	0.35	0.35	0.35	0.40	0.31	0.31	0.36	0.36	0.34	0.40	0.43	0.30	0.29	0.38
Item 22	0.86	0.87	0.86	0.86	0.80	0.79	0.89	0.61	0.61	0.62	0.67	0.56	0.51	0.62
Item 23	0.84	0.83	0.85	0.85	0.75	0.78	0.86	0.49	0.50	0.47	0.58	0.48	0.48	0.45
Item 24	0.91	0.90	0.91	0.89	0.83	0.83	0.94	0.78	0.80	0.76	0.89	0.74	0.70	0.79
Item 25	0.88	0.89	0.88	0.87	0.80	0.83	0.91	0.67	0.68	0.65	0.75	0.63	0.61	0.66
Item 26	0.91	0.91	0.91	0.89	0.82	0.84	0.94	0.79	0.80	0.78	0.85	0.73	0.70	0.80
Item 27	0.76	0.74	0.79	0.82	0.70	0.63	0.80	0.74	0.79	0.70	0.77	0.65	0.67	0.76
Item 29	0.66	0.73	0.59	0.76	0.57	0.48	0.70	0.60	0.67	0.56	0.66	0.52	0.49	0.60
Item 30	0.70	0.70	0.70	0.70	0.67	0.62	0.72	0.48	0.52	0.45	0.59	0.43	0.39	0.50
Item 31	0.54	0.54	0.54	0.56	0.47	0.44	0.57	0.55	0.58	0.51	0.52	0.49	0.51	0.55
Item 33	0.47	0.45	0.48	0.52	0.40	0.39	0.49	0.48	0.46	0.49	0.47	0.41	0.43	0.48
Item 34	0.59	0.63	0.55	0.63	0.48	0.45	0.63	0.64	0.67	0.60	0.66	0.54	0.46	0.66
Item 37	0.52	0.53	0.51	0.59	0.41	0.39	0.56	0.61	0.63	0.59	0.63	0.54	0.49	0.62
Mean	0.63	0.64	0.63	0.66	0.56	0.54	0.66	0.58	0.60	0.55	0.62	0.52	0.48	0.58
S.D.	0.19	0.18	0.19	0.16	0.18	0.19	0.19	0.11	0.12	0.11	0.12	0.12	0.13	0.11
N	23525	11692	11833	1494	2983	2862	15785							
Wtd N	2880468	1442829	1437639	101846	289984	371004	2065360							
% Answering Last Item	98%	98%	98%	97%	97%	97%	99%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

BEST COPY AVAILABLE

Appendix D-2
History/Citizenship/Geography: First Follow-up (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.83	0.84	0.81	0.74	0.69	0.72	0.87	0.66	0.71	0.60	0.69	0.62	0.55	0.65
Item 2	0.64	0.68	0.60	0.66	0.56	0.45	0.69	0.62	0.65	0.59	0.60	0.51	0.49	0.63
Item 3	0.63	0.63	0.62	0.63	0.56	0.51	0.66	0.40	0.41	0.39	0.43	0.54	0.38	0.35
Item 4	0.56	0.58	0.54	0.62	0.43	0.41	0.61	0.67	0.67	0.66	0.66	0.58	0.54	0.68
Item 5	0.68	0.64	0.72	0.77	0.66	0.57	0.70	0.58	0.61	0.59	0.56	0.56	0.61	0.58
Item 6	0.50	0.57	0.44	0.50	0.39	0.36	0.55	0.59	0.62	0.55	0.59	0.47	0.35	0.61
Item 7	0.83	0.83	0.83	0.81	0.71	0.80	0.86	0.72	0.76	0.68	0.72	0.61	0.70	0.74
Item 8	0.67	0.72	0.63	0.71	0.59	0.59	0.70	0.67	0.73	0.62	0.67	0.59	0.56	0.70
Item 9	0.52	0.55	0.50	0.67	0.61	0.45	0.52	0.46	0.45	0.46	0.42	0.41	0.46	0.48
Item 10	0.52	0.53	0.50	0.54	0.40	0.36	0.56	0.63	0.65	0.60	0.66	0.52	0.47	0.64
Item 11	0.44	0.46	0.42	0.49	0.35	0.37	0.47	0.49	0.51	0.46	0.50	0.36	0.33	0.51
Item 13	0.53	0.55	0.51	0.58	0.46	0.48	0.56	0.52	0.59	0.44	0.43	0.47	0.49	0.53
Item 14	0.80	0.81	0.79	0.83	0.71	0.69	0.84	0.62	0.68	0.57	0.66	0.57	0.53	0.62
Item 15	0.72	0.74	0.70	0.67	0.62	0.59	0.77	0.61	0.61	0.60	0.64	0.57	0.50	0.61
Item 16	0.91	0.90	0.91	0.91	0.87	0.84	0.93	0.78	0.82	0.75	0.79	0.70	0.68	0.82
Item 17	0.85	0.85	0.86	0.85	0.78	0.75	0.88	0.64	0.64	0.64	0.66	0.61	0.52	0.65
Item 18	0.28	0.31	0.26	0.34	0.23	0.22	0.30	0.46	0.46	0.46	0.53	0.41	0.28	0.49
Item 19	0.91	0.90	0.91	0.85	0.84	0.91	0.92	0.68	0.71	0.66	0.78	0.72	0.56	0.69
Item 21	0.44	0.45	0.44	0.49	0.35	0.39	0.47	0.59	0.60	0.58	0.57	0.52	0.54	0.61
Item 27	0.80	0.78	0.82	0.87	0.74	0.68	0.83	0.77	0.81	0.74	0.78	0.68	0.70	0.79
Item 29	0.74	0.82	0.67	0.83	0.64	0.62	0.78	0.69	0.74	0.66	0.67	0.64	0.55	0.70
Item 30	0.81	0.81	0.81	0.82	0.79	0.74	0.83	0.58	0.64	0.52	0.57	0.49	0.47	0.61
Item 31	0.67	0.66	0.67	0.67	0.58	0.59	0.69	0.60	0.64	0.56	0.54	0.48	0.61	0.62
Item 32	0.32	0.34	0.31	0.36	0.26	0.26	0.34	0.52	0.50	0.54	0.42	0.52	0.46	0.53
Item 33	0.60	0.59	0.61	0.64	0.53	0.54	0.62	0.55	0.55	0.54	0.46	0.50	0.47	0.55
Item 34	0.55	0.61	0.49	0.62	0.45	0.42	0.58	0.62	0.65	0.59	0.58	0.43	0.46	0.66
Item 35	0.71	0.71	0.71	0.72	0.66	0.65	0.73	0.46	0.50	0.42	0.48	0.37	0.46	0.46
Item 37	0.56	0.56	0.56	0.61	0.44	0.42	0.60	0.65	0.67	0.64	0.68	0.52	0.56	0.66
Item 38	0.45	0.51	0.40	0.49	0.43	0.35	0.47	0.44	0.52	0.34	0.34	0.37	0.31	0.46
Item 39	0.42	0.42	0.41	0.35	0.38	0.40	0.43	0.31	0.33	0.29	0.30	0.33	0.25	0.32
Mean	0.63	0.65	0.62	0.65	0.56	0.54	0.66	0.59	0.61	0.56	0.58	0.52	0.50	0.60
S.D.	0.17	0.16	0.18	0.16	0.17	0.17	0.17	0.11	0.11	0.11	0.13	0.10	0.11	0.11
N	17591	8796	8795	1096	2131	1823	12274							
Wtd N	2841095	1429618	1411477	103882	281656	361278	2047265							
% Answering Last Item	98%	98%	97%	97%	95%	95%	98%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

BEST COPY AVAILABLE

Appendix D-3
History/Citizenship/Geography: Second Follow-up (One Form Only)

Item Pool Number	Proportion Correct (P+)							R-Biserial						
	Total	Male	Female	Asian	Hispanic	Black	White	Total	Male	Female	Asian	Hispanic	Black	White
Item 1	0.89	0.90	0.88	0.84	0.78	0.82	0.92	0.67	0.73	0.62	0.64	0.65	0.63	0.66
Item 2	0.66	0.70	0.61	0.70	0.55	0.47	0.70	0.68	0.68	0.68	0.65	0.57	0.54	0.70
Item 5	0.71	0.68	0.74	0.73	0.68	0.58	0.74	0.58	0.61	0.57	0.60	0.58	0.61	0.56
Item 6	0.54	0.58	0.50	0.56	0.41	0.36	0.60	0.68	0.71	0.65	0.73	0.59	0.56	0.69
Item 8	0.76	0.80	0.72	0.79	0.69	0.68	0.79	0.69	0.72	0.67	0.70	0.66	0.60	0.72
Item 9	0.59	0.63	0.55	0.70	0.68	0.52	0.59	0.54	0.56	0.51	0.52	0.41	0.55	0.57
Item 10	0.61	0.63	0.59	0.65	0.49	0.44	0.67	0.69	0.71	0.67	0.73	0.67	0.61	0.68
Item 11	0.57	0.60	0.54	0.63	0.47	0.49	0.60	0.61	0.63	0.58	0.66	0.62	0.43	0.63
Item 12	0.41	0.44	0.37	0.39	0.31	0.34	0.43	0.44	0.41	0.47	0.49	0.35	0.31	0.46
Item 13	0.65	0.66	0.64	0.74	0.59	0.55	0.67	0.57	0.62	0.52	0.54	0.52	0.50	0.58
Item 15	0.80	0.82	0.78	0.81	0.70	0.64	0.85	0.63	0.63	0.62	0.54	0.57	0.62	0.61
Item 18	0.56	0.58	0.55	0.61	0.44	0.43	0.61	0.69	0.71	0.68	0.73	0.57	0.59	0.71
Item 19	0.96	0.96	0.96	0.94	0.92	0.94	0.97	0.56	0.61	0.50	0.61	0.55	0.50	0.52
Item 20	0.43	0.45	0.42	0.46	0.38	0.31	0.46	0.53	0.54	0.52	0.56	0.41	0.38	0.56
Item 21	0.59	0.58	0.59	0.63	0.48	0.51	0.62	0.71	0.71	0.72	0.62	0.66	0.71	0.73
Item 27	0.91	0.90	0.92	0.95	0.87	0.86	0.92	0.74	0.78	0.71	0.72	0.67	0.69	0.76
Item 28	0.52	0.55	0.49	0.56	0.45	0.43	0.55	0.49	0.52	0.45	0.48	0.38	0.33	0.51
Item 31	0.78	0.79	0.78	0.77	0.67	0.68	0.82	0.66	0.69	0.64	0.73	0.62	0.64	0.66
Item 32	0.43	0.45	0.42	0.45	0.37	0.36	0.46	0.55	0.54	0.56	0.53	0.53	0.53	0.55
Item 33	0.72	0.71	0.74	0.73	0.64	0.69	0.75	0.60	0.62	0.59	0.56	0.63	0.48	0.61
Item 36	0.25	0.27	0.24	0.27	0.23	0.22	0.26	0.28	0.34	0.21	0.33	0.25	0.15	0.30
Item 37	0.68	0.68	0.68	0.73	0.57	0.58	0.72	0.68	0.71	0.64	0.64	0.64	0.63	0.67
Item 40	0.63	0.68	0.57	0.73	0.58	0.54	0.65	0.60	0.63	0.56	0.71	0.55	0.50	0.62
Item 41	0.70	0.67	0.73	0.73	0.66	0.63	0.72	0.46	0.46	0.48	0.44	0.46	0.44	0.44
Item 42	0.56	0.56	0.55	0.57	0.50	0.48	0.58	0.60	0.61	0.59	0.75	0.52	0.45	0.62
Item 43	0.64	0.66	0.62	0.65	0.50	0.44	0.70	0.65	0.70	0.59	0.66	0.59	0.60	0.63
Item 44	0.55	0.57	0.52	0.58	0.48	0.46	0.57	0.50	0.54	0.46	0.51	0.42	0.49	0.50
Item 45	0.29	0.29	0.29	0.35	0.22	0.21	0.31	0.48	0.45	0.52	0.60	0.33	0.25	0.51
Item 46	0.35	0.37	0.32	0.39	0.29	0.26	0.37	0.42	0.40	0.43	0.36	0.43	0.28	0.44
Item 47	0.20	0.22	0.19	0.26	0.21	0.17	0.20	0.30	0.30	0.28	0.41	0.30	0.13	0.32
Mean	0.60	0.61	0.58	0.63	0.53	0.50	0.63	0.58	0.60	0.56	0.59	0.52	0.49	0.58
S.D.	0.18	0.18	0.19	0.18	0.18	0.18	0.19	0.11	0.12	0.11	0.11	0.12	0.15	0.11
N	14063	7029	7034	931	1732	1377	9830							
Wtd N	2253399	1155060	1098339	89668	232262	286537	1611023							
% Answering Last Item	97%	97%	97%	95%	93%	95%	98%							

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix E: Test Item Map

Appendix E-1
Test Item Map
Reading

Answer Key	# Valid Choices	88	Item Number in Booklet 90L	90H		92L	92H		IRT Parameters A	B	C
1 3(C)	5		1	1			1		1.18120	-2.51737	0.00000
2 2(B)	5		2	2			2		0.92613	-1.95897	0.00000
3 4(D)	5		3	3			3		0.96886	-1.72667	0.00000
4 5(E)	5		4	4			4		0.80503	-0.82988	0.00000
5 3(C)	5		5	5			5		1.12384	-0.36093	0.19648
6 1(A)	5				1				0.84073	0.72554	0.31302
7 1(A)	5				2				0.85544	0.91442	0.26454
8 5(E)	5				3				0.86801	0.78061	0.19714
9 5(E)	5				4				1.01054	0.06088	0.06813
10 3(C)	5				5				0.82278	0.75733	0.21344
11 5(E)	5				6				1.10353	-0.76371	0.00000
12 2(B)	5				7				0.78865	0.24552	0.03371
13 5(E)	5				8				0.98421	-0.42050	0.00000
14 1(A)	5						13		1.76071	0.88232	0.16581
15 4(D)	5		6	6	9		14		0.89603	-0.81761	0.11054
16 4(D)	5		7	7	10		15		0.84671	0.06466	0.08756
17 3(C)	5		8	8	11		16		0.89737	-0.43866	0.07115
18 3(C)	4		9	9					0.74775	-0.46042	0.26892
19 4(D)	4		10	10	12		6	5	0.32190	0.21636	0.00000
20 1(A)	4		11		13				0.69730	-0.73147	0.06883
21 1(A)	4			11					0.72059	-1.44086	0.00000
22 4(D)	4		12	12	14		7	6	1.16762	-1.03718	0.14815
23 3(C)	4		13	13	15		8	7	1.29257	0.07275	0.32389
24 4(D)	4		14	14	16		9	8	1.32902	-0.17197	0.19616
25 4(D)	4							4	0.59540	1.53796	0.17597
26 3(C)	4							3	0.51022	-0.45631	0.00000
27 2(B)	4							1	0.59259	-1.69826	0.00000
28 2(B)	4							2	0.93951	-0.66506	0.04337
29 4(D)	5				17				0.68568	0.98921	0.19949
30 3(C)	5				18				0.55649	0.30714	0.20377
31 2(B)	5				19				0.88084	-0.62245	0.00000
32 1(A)	5				20				0.52940	0.97253	0.06243
33 4(D)	5				21				0.45735	1.95894	0.13639
34 4(D)	5							13	0.57560	0.21277	0.00000
35 4(D)	5							14	1.11779	1.96346	0.18166
36 5(E)	5							15	0.96984	1.18825	0.15996
37 2(B)	5							16	1.19692	1.59917	0.20184
38 4(D)	4		15	15			10		0.99102	-0.28401	0.08331
39 1(A)	4		16	16			11		1.25847	-1.23530	0.24453
40 1(A)	4		17	17					1.62555	-0.09671	0.26114
41 2(B)	4		18	18			12		0.63049	-0.31581	0.16434
42 3(C)	4		19	19					1.07807	-0.66149	0.20750
43 2(B)	4		20	20					1.04897	-0.81284	0.32658
44 3(C)	4		21	21					1.23138	-0.35399	0.31870
45 2(B)	4						17	17	1.14014	-0.07623	0.45227
46 3(C)	4						18	18	1.25230	1.06442	0.35039
47 2(B)	4						19		1.14844	-0.68559	0.31178
48 1(A)	4						20	20	0.59287	1.07591	0.17999
49 3(C)	4						21	21	0.83143	0.97458	0.22774
50 3(C)	4							9	0.81723	0.06436	0.21675
51 4(D)	4							10	0.52141	1.25622	0.10153
52 4(D)	4							11	0.61980	1.73954	0.17764
53 1(A)	4							12	0.49945	1.75052	0.15205
54 4(D)	4							19	1.02749	2.34088	0.19858

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix E-2
Test Item Map
Math

	Answer Key	# Valid Choices	Item Number in Booklet							IRT Parameters		
			88	90L	90M	90H	92L	92M	92H	A	B	C
1	4 (D)	4	28	29	23	19	30	19		0.68181	-0.87241	0.11087
2	2 (B)	4					26			0.81955	-0.76121	0.17258
3	4 (D)	5	21	22		16	22			0.59218	-1.64137	0.00000
4	1 (A)	4		40			17			0.80777	-2.94873	0.06710
5	4 (D)	5	29	30	24	20				0.79283	-0.66171	0.08814
6	3 (C)	4	31	32	26		28			0.83407	-1.08544	0.09471
7	2 (B)	5	25	26			24			0.89889	-1.10120	0.15730
8	2 (B)	4	34	34	28	23	29			1.01292	-0.47088	0.24387
9	3 (C)	4	26	27	22	18	23	17		1.12383	-0.46246	0.35119
10	3 (C)	4	32	33						0.87113	-0.74347	0.35651
11	2 (B)	4	5	3	5	4	9	4		1.29364	-0.53688	0.21087
12	4 (D)	4	4	2	4	3	10	6		1.19470	-0.33819	0.20949
13	2 (B)	4	9	4	9	8	11			1.01044	0.09795	0.23418
14	1 (A)	4		7			2			0.71930	-2.22133	0.00000
15	4 (D)	4	7		7	6				1.07586	-0.11721	0.11326
16	3 (C)	4	12	11	12	11				0.79942	-0.40340	0.05706
17	1 (A)	4	2		2	1				0.60453	-0.53500	0.07134
18	1 (A)	4	3		3	2				0.92699	0.95693	0.40262
19	1 (A)	4		8						1.24943	0.01075	0.19848
20	3 (C)	4		9						1.40404	-0.05373	0.21384
21	1 (A)	4		6			8			0.56981	-0.92211	0.19984
22	2 (B)	4	13	12	13	12	12	9		0.88153	-0.60426	0.09364
23	4 (D)	4	10	5	10	9	15	11		0.96547	0.04512	0.17120
24	2 (B)	4	6		6	5		12	2	1.00754	0.45108	0.30110
25	2 (B)	4	8		8	7		13	3	0.68957	0.27051	0.09071
26	1 (A)	4	11	10	11	10	16	10	1	0.82091	0.11529	0.11306
27	1 (A)	4							4	0.98903	2.29678	0.11834
28	1 (A)	4	14	13	14	13	14	7		1.06022	-0.32865	0.14891
29	1 (A)	4	15	14		14	7			0.99843	-0.61601	0.43884
30	2 (B)	4	16	15	15		3	3		0.54766	-2.19425	0.00000
31	2 (B)	4	17	16	16		5	5		0.54485	-0.76427	0.38465
32	2 (B)	4	18	17	17	15	13	8		1.15688	-0.26050	0.21053
33	2 (B)	4	19	18	18		1	1		0.68679	-2.21344	0.03540
34	3 (C)	5	33		27	22	34	24		0.54566	0.93151	0.32992
35	2 (B)	4	24	25	21	17	27	16		0.57035	-1.18917	0.02352
36	4 (D)	4	30	31	25	21	31	21	8	0.58607	-0.41898	0.13473
37	2 (B)	4	39	38	33	28	40	23	10	1.30207	0.06324	0.12511
38	4 (D)	4	37		31	26				0.83285	-0.59678	0.00000
39	4 (D)	5	40	39	34	29	33	18	6	1.08731	-0.19037	0.11735
40	2 (B)	4	38	37	32	27		27	13	1.36826	1.29155	0.34865
41	2 (B)	4						34	26	1.14429	2.25687	0.25864
42	5 (E)	5							29	0.69035	1.26821	0.00000
43	3 (C)	4				30		38	32	0.64398	2.41658	0.12428
44	4 (D)	4	36	36	30	25	36	20	7	0.92334	0.01612	0.12642
45	3 (C)	5					38	36	22	0.60561	2.27172	0.22935
46	3 (C)	4				31			23	1.12318	1.40632	0.22014
47	3 (C)	4				32			19	0.67679	2.00317	0.25383
48	3 (C)	4							28	1.48766	2.12629	0.19798
49	2 (B)	5				33			9	2.14550	1.07065	0.34743
50	3 (C)	4	35	35	29	24	25	22		0.60185	-0.22727	0.26618
51	3 (C)	3			35	34		25	12	0.83282	0.13847	0.10066
52	1 (A)	4				35			20	1.36009	1.15455	0.06559
53	4 (D)	5			36	36				0.59898	-0.46164	0.04239
54	3 (C)	5			37	37		28	11	1.41513	1.01649	0.24226
55	1 (A)	5			38	38		30	18	0.95161	1.01715	0.20330

Appendix E-2
Test Item Map
Math (Continued)

	Answer Key	# Valid Choices	Item Number in Booklet							IRT Parameters		
			88	90L	90M	90H	92L	92M	92H	A	B	C
56	3 (C)	5			39	39		32	24	0.73958	1.25686	0.16181
57	1 (A)	5			40	40		31	17	0.85972	0.85092	0.10950
58	5 (E)	5						40	40	1.33843	2.81896	0.04093
59	2 (B)	5						39	37	1.31305	2.77701	0.15386
60	1 (A)	4	1	1	1		6	2		1.13553	-1.31660	0.20392
61	4 (D)	4	20	21	19		18	14		0.75484	-2.25518	0.00000
62	1 (A)	4	22	23			19			0.90953	-1.58401	0.00000
63	3 (C)	4	23	24	20		20	15		0.41684	-1.58628	0.00000
64	3 (C)	4	27	28			32			1.55719	-0.74660	0.16430
65	2 (B)	4		19						1.11627	-0.00395	0.16357
66	3 (C)	4		20			4			0.86183	-1.94097	0.00000
67	5 (E)	5					21		5	0.52694	-1.59965	0.00000
68	5 (E)	5					35		15	1.14276	0.46401	0.08410
69	4 (D)	4					37	35	21	0.54005	1.35221	0.18907
70	4 (D)	5					39	26	14	0.83555	0.50640	0.09662
71	1 (A)	5						29	16	0.68308	2.47157	0.40168
72	3 (C)	5						33	25	0.98551	2.01246	0.29597
73	5 (E)	5						37	27	0.96775	1.59789	0.08675
74	4 (D)	5							30	0.68921	2.77731	0.22115
75	1 (A)	4							31	1.01358	1.82906	0.14133
76	4 (D)	4							33	1.59430	2.11449	0.12061
77	3 (C)	5							34	1.31935	2.29660	0.14979
78	1 (A)	4							35	1.07980	3.20302	0.11385
79	4 (D)	5							36	0.89043	2.91767	0.12718
80	5 (E)	5							38	1.29152	2.56220	0.05966
81	4 (D)	5							39	1.49669	2.66925	0.11299

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix E-3
Test Item Map
Science

	Answer Key	# Valid Choices	Item Number in Booklet 88	90	92	IRT Parameters A	B	C
1	3 (C)	4	1			1.16608	-0.67228	0.37787
2	5 (E)	5	2			0.59777	-1.93399	0.13876
3	1 (A)	4	3	2		0.69979	-0.57676	0.33921
4	3 (C)	4	4	3	5	0.66591	-0.62182	0.31695
5	5 (E)	5	5	4	2	1.09400	-1.36000	0.00000
6	5 (E)	5	6	5	1	1.04363	-1.55512	0.00002
7	1 (A)	4	7			0.52146	-1.29720	0.00000
8	1 (A)	4	8			0.62419	-0.25581	0.25386
9	2 (B)	5	9			0.53319	-1.36224	0.00001
10	3 (C)	4	10	1	8	1.10474	0.00281	0.30008
11	3 (C)	4	11			0.43784	0.20647	0.19275
12	3 (C)	5	12	6	6	0.85169	-0.65205	0.27561
13	4 (D)	4	13			0.60663	-1.75538	0.00001
14	3 (C)	5	14	7	3	1.23878	-0.41510	0.19739
15	1 (A)	4	15	8	15	0.40637	-0.28296	0.00001
16	3 (C)	4	16	9	18	0.95246	0.47833	0.33145
17	2 (B)	4	17	10	7	1.28611	0.12036	0.25544
18	2 (B)	4	18	11	9	0.97920	0.00387	0.22460
19	3 (C)	4	19	12	14	1.01363	0.24806	0.24407
20	2 (B)	4	20	13		1.15653	0.74217	0.33252
21	3 (C)	4	21	14		0.96782	0.61829	0.31361
22	4 (D)	4	22	15	16	0.67782	0.90750	0.25591
23	3 (C)	4	23	16		1.43791	1.05388	0.38865
24	1 (A)	5	24	17	20	0.62227	0.20736	0.00001
25	4 (D)	5	25	18		0.64546	1.18072	0.09492
26	3 (C)	4		20	19	0.88578	0.01877	0.16607
27	4 (D)	4		19	21	1.46803	0.99365	0.13903
28	1 (A)	4			4	0.70864	-0.36201	0.34331
29	1 (A)	4		21	12	1.09783	0.18743	0.17761
30	2 (B)	5		22	13	0.80216	0.27046	0.21798
31	4 (D)	4			10	0.37842	-0.57463	0.00001
32	1 (A)	4		23	22	1.43394	0.96323	0.12356
33	4 (D)	4		24	11	0.80165	-0.32345	0.10520
34	1 (A)	4		25		0.32691	0.10811	0.00000
35	1 (A)	4			17	1.04588	0.81089	0.21361
36	2 (B)	4			23	0.71678	1.76348	0.32502
37	1 (A)	4			24	0.81268	2.18077	0.23181
38	4 (D)	4			25	1.54588	2.40482	0.10371

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix E-4
Test Item Map
History/Citizenship/Geography

	Answer Key	# Valid Choices	Item Number in Booklet 88	90	92	IRT Parameters A	B	C
1	3(C)	4	4	1	2	0.98219	-1.25256	0.21137
2	3(C)	4	26	2	14	1.12623	0.00140	0.28845
3	2(B)	4		3		0.29554	-1.37111	0.00000
4	1(A)	4	22	4		1.45953	-0.02180	0.26657
5	1(A)	4	12	5	6	0.57016	-0.93455	0.02822
6	2(B)	4	28	6	18	1.52760	0.44390	0.27880
7	4(D)	4	2	7		1.10537	-1.33515	0.26274
8	4(D)	4	13	8	3	1.36141	-0.26818	0.32572
9	3(C)	4	14	9	10	0.75018	0.47592	0.25624
10	5(E)	5	15	10	12	1.02945	0.02726	0.18382
11	2(B)	5	16	11	13	1.24221	0.56911	0.29637
12	2(B)	4			26	1.48652	1.48763	0.29832
13	3(C)	4	23	12	11	0.93498	0.28607	0.29308
14	2(B)	4	18	13		0.87587	-1.26965	0.33294
15	4(D)	4	20	14	7	0.71144	-1.13364	0.08806
16	3(C)	4	3	15		2.03444	-1.52077	0.46357
17	2(B)	4	1	16		1.07288	-1.08690	0.48813
18	2(B)	4	30	17	25	1.88350	0.75941	0.19735
19	1(A)	4	17	18	1	1.00430	-1.84445	0.27435
20	3(C)	4			22	1.30349	1.25515	0.26184
21	1(A)	4	29	19	16	1.35758	0.50549	0.23433
22	1(A)	2	5			0.96925	-1.92663	0.23751
23	1(A)	2	6			0.52152	-2.69376	0.00000
24	2(B)	2	7			1.64167	-2.11534	0.00000
25	1(A)	2	8			1.03994	-2.19188	0.00000
26	2(B)	2	9			1.75480	-2.12320	0.00000
27	4(D)	5	19	20	4	1.49480	-1.14670	0.24233
28	2(B)	4			21	0.88606	0.99954	0.29325
29	2(B)	4	21	21		1.20516	-0.62570	0.35219
30	3(C)	4	10	22		1.10922	-0.44457	0.51625
31	4(D)	4	24	23	5	0.84672	-0.60389	0.15013
32	1(A)	4		24	23	0.63192	0.82388	0.07269
33	2(B)	4	25	25	9	0.76584	-0.22218	0.21016
34	2(B)	4	11	26		1.59962	-0.06140	0.30746
35	2(B)	4		27		0.44765	-1.46990	0.00168
36	1(A)	4			29	1.25594	2.25819	0.20646
37	1(A)	4	27	28	15	0.90837	-0.30759	0.13674
38	4(D)	4		29		0.93793	0.77969	0.28098
39	2(B)	4		30		0.68855	1.62702	0.31263
40	3(C)	4			17	1.15943	0.48314	0.32292
41	1(A)	4			8	0.41296	-1.05935	0.00000
42	3(C)	4			19	1.32067	0.75449	0.30523
43	4(D)	4			20	0.97527	0.14559	0.21349
44	2(B)	4			24	0.70172	0.80714	0.25314
45	3(C)	4			27	1.11145	1.64311	0.15251
46	2(B)	4			28	1.02496	1.71842	0.22389
47	1(A)	4			30	1.28831	2.25424	0.15843

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix F: Invariance of Item Parameters Across Years

Appendix F-1
Invariance of Item Parameters Across Years
Reading Test

Item #	Used In Test Forms			Number of Responses			Proportion Correct for Item Respondents: Raw Item Responses			Proportion Correct for Item Respondents: IRT Estimates			Deviation (Actual-Predicted)		
				BY	F1	F2	BY	F1	F2	BY	F1	F2	BY	F1	F2
1	BY	F1L	F2L	23605	9100	7071	0.95	0.92	0.94	0.95	0.93	0.94	0.00	-0.01	-0.01
2	BY	F1L	F2L	23577	9086	7067	0.86	0.80	0.82	0.85	0.80	0.84	0.01	0.00	-0.02
3	BY	F1L	F2L	23577	9088	7065	0.82	0.77	0.80	0.81	0.76	0.80	0.01	0.01	0.00
4	BY	F1L	F2L	23536	9077	7060	0.58	0.50	0.57	0.58	0.50	0.56	0.00	0.00	0.01
5	BY	F1L	F2L	23449	9044	7033	0.56	0.46	0.57	0.56	0.47	0.54	0.00	-0.01	0.03
6		F1 H		0	8701	0	NA	0.63	NA	NA	0.63	NA	NA	0.00	NA
7		F1 H		0	8664	0	NA	0.55	NA	NA	0.56	NA	NA	-0.01	NA
8		F1 H		0	8666	0	NA	0.56	NA	NA	0.55	NA	NA	0.00	NA
9		F1 H		0	8688	0	NA	0.66	NA	NA	0.69	NA	NA	-0.03	NA
10		F1 H		0	8673	0	NA	0.58	NA	NA	0.57	NA	NA	0.01	NA
11		F1 H		0	8671	0	NA	0.85	NA	NA	0.87	NA	NA	-0.02	NA
12		F1 H		0	8665	0	NA	0.60	NA	NA	0.61	NA	NA	0.00	NA
13		F1 H		0	8657	0	NA	0.76	NA	NA	0.78	NA	NA	-0.02	NA
14			F2L	0	0	6977	NA	NA	0.25	NA	NA	0.23	NA	NA	0.02
15	BY	F1LH	F2L	23592	17811	6977	0.60	0.69	0.59	0.63	0.70	0.61	-0.02	-0.01	-0.02
16	BY	F1LH	F2L	23552	17770	6937	0.41	0.50	0.37	0.40	0.49	0.38	0.01	0.00	-0.01
17	BY	F1LH	F2L	23545	17796	6940	0.49	0.62	0.46	0.51	0.60	0.49	-0.02	0.02	-0.03
18	BY	F1L		22528	8636	0	0.64	0.57	NA	0.63	0.57	NA	0.01	0.00	NA
19	BY	F1LH	F2LH	22417	17734	14191	0.42	0.44	0.46	0.41	0.45	0.49	0.01	-0.02	-0.03
20	BY	F1 H		23438	8684	0	0.59	0.77	NA	0.58	0.81	NA	0.01	-0.04	NA
21		F1L		0	9031	0	NA	0.65	NA	NA	0.65	NA	NA	-0.01	NA
22	BY	F1LH	F2LH	23444	17712	14185	0.71	0.76	0.78	0.71	0.77	0.82	0.00	-0.01	-0.04
23	BY	F1LH	F2LH	23371	17670	14162	0.51	0.63	0.68	0.53	0.62	0.67	-0.03	0.01	0.01
24	BY	F1LH	F2LH	23294	17611	14141	0.48	0.61	0.67	0.50	0.60	0.66	-0.02	0.01	0.00
25			F2 H	0	0	7120	NA	NA	0.47	NA	NA	0.47	NA	NA	0.00
26			F2 H	0	0	7143	NA	NA	0.70	NA	NA	0.75	NA	NA	-0.05
27			F2 H	0	0	7137	NA	NA	0.90	NA	NA	0.92	NA	NA	-0.02
28			F2 H	0	0	7145	NA	NA	0.87	NA	NA	0.89	NA	NA	-0.02
29		F1 H		0	8496	0	NA	0.52	NA	NA	0.52	NA	NA	0.00	NA
30		F1 H		0	8597	0	NA	0.64	NA	NA	0.65	NA	NA	-0.01	NA
31		F1 H		0	8570	0	NA	0.79	NA	NA	0.81	NA	NA	-0.03	NA
32		F1 H		0	8560	0	NA	0.45	NA	NA	0.46	NA	NA	0.00	NA
33		F1 H		0	8542	0	NA	0.36	NA	NA	0.37	NA	NA	-0.01	NA
34			F2 H	0	0	6989	NA	NA	0.60	NA	NA	0.65	NA	NA	-0.04
35			F2 H	0	0	6972	NA	NA	0.32	NA	NA	0.33	NA	NA	-0.01
36			F2 H	0	0	7007	NA	NA	0.51	NA	NA	0.51	NA	NA	0.01
37			F2 H	0	0	7003	NA	NA	0.43	NA	NA	0.41	NA	NA	0.02
38	BY	F1L	F2L	23251	8927	7009	0.47	0.38	0.49	0.48	0.39	0.46	-0.01	-0.01	0.03
39	BY	F1L	F2L	23142	8884	7006	0.77	0.71	0.79	0.79	0.73	0.78	-0.02	-0.02	0.01
40	BY	F1L		23046	8826	0	0.54	0.40	NA	0.52	0.43	NA	0.02	-0.03	NA

Appendix F-1
Invariance of Item Parameters Across Years
Reading Test (Continued)

Item #	Used In Test Forms			Number of Responses			Proportion Correct for Item Respondents: Raw Item Responses			Proportion Correct for Item Respondents: IRT Estimates			Deviation (Actual-Predicted)		
				BY	F1	F2	BY	F1	F2	BY	F1	F2	BY	F1	F2
41	BY	F1L	F2L	22961	8809	6994	0.54	0.46	0.55	0.54	0.48	0.53	0.00	-0.02	0.02
42	BY	F1L		22765	8714	0	0.63	0.55	NA	0.64	0.56	NA	-0.01	-0.01	NA
43	BY	F1L		22714	8684	0	0.71	0.67	NA	0.72	0.66	NA	-0.02	0.01	NA
44	BY	F1L		22638	8651	0	0.62	0.55	NA	0.62	0.55	NA	0.00	0.01	NA
45			F2LH	0	0	13656	NA	NA	0.75	NA	NA	0.76	NA	NA	0.00
46			F2LH	0	0	13487	NA	NA	0.52	NA	NA	0.51	NA	NA	0.01
47			F2L	0	0	6672	NA	NA	0.69	NA	NA	0.68	NA	NA	0.01
48			F2LH	0	0	13282	NA	NA	0.45	NA	NA	0.44	NA	NA	0.01
49			F2LH	0	0	13186	NA	NA	0.46	NA	NA	0.46	NA	NA	0.01
50			F2 H	0	0	7097	NA	NA	0.78	NA	NA	0.78	NA	NA	-0.01
51			F2 H	0	0	7118	NA	NA	0.49	NA	NA	0.49	NA	NA	0.00
52			F2 H	0	0	7055	NA	NA	0.44	NA	NA	0.44	NA	NA	0.00
53			F2 H	0	0	7080	NA	NA	0.44	NA	NA	0.44	NA	NA	0.00
54			F2 H	0	0	6682	NA	NA	0.30	NA	NA	0.30	NA	NA	0.00
									Sum of Deviations for All Items				0.08	0.08	0.20

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix F-2
Invariance of Item Parameters Across Years
Math Test

Item #	Used In Test Forms			Number of Responses			Proportion Correct for Item Respondents: Raw Item Responses			Proportion Correct for Item Respondents: IRT Estimates			Deviation (Actual-Predicted)		
				BY	F1	F2	BY	F1	F2	BY	F1	F2	BY	F1	F2
1	BY	F1LMH	F2LM	23407	17709	10226	0.57	0.69	0.70	0.58	0.69	0.69	-0.01	-0.01	0.01
2			F2L	0	0	2349	NA	NA	0.49	NA	NA	0.49	NA	NA	0.00
3	BY	F1L H	F2L	23247	7937	2507	0.70	0.75	0.59	0.68	0.77	0.61	0.02	-0.02	-0.02
4		F1L	F2L	0	3125	2533	NA	0.85	0.90	NA	0.88	0.91	NA	-0.03	-0.01
5	BY	F1LMH		23191	17578	0	0.53	0.65	NA	0.53	0.66	NA	0.00	-0.01	NA
6	BY	F1LM	F2L	22988	12821	2517	0.61	0.69	0.59	0.62	0.68	0.53	-0.02	0.01	0.06
7	BY	F1L	F2L	23439	3182	2536	0.66	0.48	0.57	0.65	0.50	0.57	0.00	-0.02	0.01
8	BY	F1LMH	F2L	22651	17377	2390	0.53	0.73	0.48	0.56	0.69	0.45	-0.03	0.04	0.03
9	BY	F1LMH	F2LM	23162	17503	10073	0.64	0.75	0.72	0.62	0.74	0.73	0.01	0.01	-0.01
10	BY	F1L		22965	3088	0	0.68	0.53	NA	0.68	0.55	NA	0.00	-0.02	NA
11	BY	F1LMH	F2LM	22889	17517	10158	0.53	0.71	0.70	0.56	0.70	0.69	-0.03	0.01	0.01
12	BY	F1LMH	F2LM	23342	17629	10197	0.50	0.65	0.66	0.51	0.65	0.64	-0.01	0.00	0.02
13	BY	F1LMH	F2L	23351	17536	2508	0.44	0.58	0.35	0.45	0.58	0.34	0.00	0.00	0.01
14		F1L	F2L	0	3084	2511	NA	0.74	0.81	NA	0.71	0.77	NA	0.03	0.04
15	BY	F1 MH		23244	14417	0	0.42	0.62	NA	0.40	0.64	NA	0.02	-0.02	NA
16	BY	F1LMH		23372	17657	0	0.45	0.58	NA	0.45	0.59	NA	0.00	-0.01	NA
17	BY	F1 MH		23414	14510	0	0.50	0.64	NA	0.49	0.67	NA	0.01	-0.03	NA
18	BY	F1 MH		22959	14360	0	0.49	0.58	NA	0.48	0.59	NA	0.01	-0.01	NA
19		F1L		0	3151	0	NA	0.27	NA	NA	0.27	NA	NA	0.00	NA
20		F1L		0	3118	0	NA	0.28	NA	NA	0.28	NA	NA	0.00	NA
21		F1L	F2L	0	3107	2501	NA	0.56	0.52	NA	0.52	0.56	NA	0.04	-0.04
22	BY	F1LMH	F2LM	23113	17475	10096	0.54	0.64	0.64	0.51	0.65	0.65	0.02	-0.01	-0.01
23	BY	F1LMH	F2LM	23365	17628	10151	0.41	0.55	0.55	0.42	0.55	0.53	0.00	-0.01	0.02
24	BY	F1 MH	F2 MH	23064	14421	11553	0.46	0.60	0.65	0.45	0.60	0.68	0.01	0.00	-0.03
25	BY	F1 MH	F2 MH	23494	14508	11631	0.37	0.51	0.57	0.35	0.53	0.61	0.02	-0.02	-0.04
26	BY	F1LMH	F2LMH	23237	17556	14069	0.36	0.53	0.58	0.38	0.51	0.59	-0.01	0.02	-0.01
27			F2 H	0	0	3917	NA	NA	0.41	NA	NA	0.41	NA	NA	0.00
28	BY	F1LMH	F2LM	23195	17548	10151	0.51	0.62	0.58	0.48	0.62	0.61	0.03	-0.01	-0.03
29	BY	F1L H	F2L	23004	7877	2503	0.73	0.81	0.57	0.70	0.80	0.62	0.03	0.01	-0.05
30	BY	F1LM	F2LM	23430	12886	10237	0.80	0.79	0.84	0.76	0.79	0.83	0.04	0.00	0.00
31	BY	F1LM	F2LM	23397	12861	10205	0.71	0.73	0.75	0.69	0.72	0.76	0.01	0.01	-0.01
32	BY	F1LMH	F2LM	23296	17683	10223	0.52	0.62	0.61	0.49	0.63	0.62	0.03	-0.01	-0.01
33	BY	F1LM	F2LM	23113	12906	10254	0.81	0.85	0.86	0.81	0.84	0.88	0.01	0.01	-0.02
34	BY	F1 MH	F2LM	23264	14456	10157	0.47	0.56	0.55	0.47	0.57	0.52	0.00	-0.01	0.03
35	BY	F1LMH	F2LM	23348	17387	10164	0.60	0.70	0.68	0.60	0.70	0.70	0.00	0.00	-0.02
36	BY	F1LMH	F2LMH	22812	17477	13961	0.54	0.62	0.67	0.51	0.62	0.68	0.03	0.00	-0.00
37	BY	F1LMH	F2LMH	22977	17503	14076	0.39	0.53	0.56	0.36	0.52	0.61	0.03	0.00	-0.05
38	BY	F1 MH		23164	14471	0	0.45	0.69	NA	0.46	0.69	NA	-0.01	0.00	NA
39	BY	F1LMH	F2LMH	22515	17275	14032	0.29	0.65	0.72	0.42	0.58	0.66	-0.14	0.07	0.06
40	BY	F1LMH	F2 MH	22837	17465	11483	0.43	0.45	0.52	0.39	0.45	0.55	0.04	0.00	-0.03
41			F2 MH	0	0	11202	NA	NA	0.35	NA	NA	0.35	NA	NA	0.00

Appendix F-2
Invariance of Item Parameters Across Years
Math Test (Continued)

Item #	Used In Test Forms			Number of Responses			Proportion Correct for Item Respondents: Raw Item Responses			Proportion Correct for Item Respondents: IRT Estimates			Deviation (Actual-Predicted)		
				BY	F1	F2	BY	F1	F2	BY	F1	F2	BY	F1	F2
42			F2 H	0	0	3699	NA	NA	0.55	NA	NA	0.64	NA	NA	-0.08
43		F1 H	F2 MH	0	4783	11255	NA	0.31	0.27	NA	0.33	0.27	NA	-0.02	0.00
44	BY	F1LMH	F2LMH	23057	17554	14122	0.41	0.53	0.61	0.39	0.53	0.61	0.02	0.00	0.00
45			F2LMH	0	0	13921	NA	NA	0.36	NA	NA	0.36	NA	NA	0.00
46		F1 H	F2 H	0	4771	3910	NA	0.56	0.72	NA	0.56	0.72	NA	-0.01	0.00
47		F1 H	F2 H	0	4716	3794	NA	0.46	0.62	NA	0.49	0.58	NA	-0.02	0.03
48			F2 H	0	0	3819	NA	NA	0.48	NA	NA	0.48	NA	NA	0.00
49		F1 H	F2 H	0	4717	3940	NA	0.68	0.91	NA	0.73	0.88	NA	-0.06	0.03
50	BY	F1LMH	F2LM	22921	17512	10102	0.58	0.65	0.62	0.56	0.65	0.64	0.02	0.00	-0.01
51		F1 MH	F2 MH	0	14264	11492	NA	0.54	0.68	NA	0.56	0.65	NA	-0.02	0.03
52		F1 H	F2 H	0	4596	3880	NA	0.55	0.78	NA	0.57	0.76	NA	-0.02	0.02
53		F1 MH		0	14326	0	NA	0.64	NA	NA	0.65	NA	NA	-0.01	NA
54		F1 MH	F2 MH	0	14057	11044	NA	0.47	0.53	NA	0.44	0.53	NA	0.02	0.00
55		F1 MH	F2 MH	0	13993	11202	NA	0.46	0.51	NA	0.44	0.52	NA	0.02	-0.01
56		F1 MH	F2 MH	0	14000	11312	NA	0.40	0.46	NA	0.39	0.46	NA	0.01	0.00
57		F1 MH	F2 MH	0	13966	11319	NA	0.41	0.51	NA	0.41	0.50	NA	0.00	0.01
58			F2 MH	0	0	10307	NA	NA	0.10	NA	NA	0.09	NA	NA	0.01
59			F2 MH	0	0	10170	NA	NA	0.21	NA	NA	0.20	NA	NA	0.01
60	BY	F1LM	F2LM	22701	12538	10221	0.74	0.74	0.85	0.73	0.78	0.84	0.01	-0.04	0.01
61	BY	F1LM	F2LM	21387	12655	10128	0.88	0.88	0.92	0.82	0.85	0.89	0.05	0.03	0.02
62	BY	F1L	F2L	23300	3151	2523	0.69	0.56	0.67	0.71	0.56	0.64	-0.02	0.00	0.03
63	BY	F1LM	F2LM	23240	12591	10064	0.66	0.69	0.71	0.63	0.66	0.71	0.03	0.02	0.00
64	BY	F1L	F2L	23207	3111	2475	0.61	0.34	0.33	0.58	0.36	0.43	0.03	-0.02	-0.10
65		F1L		0	3114	0	NA	0.24	NA	NA	0.25	NA	NA	-0.01	NA
66		F1L	F2L	0	3157	2538	NA	0.68	0.80	NA	0.66	0.74	NA	0.01	0.07
67			F2L H	0	0	6440	NA	NA	0.80	NA	NA	0.80	NA	NA	0.00
68			F2L H	0	0	6396	NA	NA	0.58	NA	NA	0.58	NA	NA	0.00
69			F2LMH	0	0	13981	NA	NA	0.45	NA	NA	0.44	NA	NA	0.01
70			F2LMH	0	0	13899	NA	NA	0.51	NA	NA	0.50	NA	NA	0.00
71			F2 MH	0	0	11276	NA	NA	0.51	NA	NA	0.49	NA	NA	0.02
72			F2 MH	0	0	10473	NA	NA	0.44	NA	NA	0.42	NA	NA	0.02
73			F2 MH	0	0	10865	NA	NA	0.34	NA	NA	0.33	NA	NA	0.01
74			F2 H	0	0	3631	NA	NA	0.44	NA	NA	0.42	NA	NA	0.02
75			F2 H	0	0	3828	NA	NA	0.55	NA	NA	0.56	NA	NA	-0.01
76			F2 H	0	0	3442	NA	NA	0.46	NA	NA	0.44	NA	NA	0.03
77			F2 H	0	0	3492	NA	NA	0.40	NA	NA	0.41	NA	NA	-0.01
78			F2 H	0	0	3021	NA	NA	0.20	NA	NA	0.20	NA	NA	-0.01
79			F2 H	0	0	3540	NA	NA	0.31	NA	NA	0.29	NA	NA	0.02
80			F2 H	0	0	3166	NA	NA	0.26	NA	NA	0.26	NA	NA	-0.01
81			F2 H	0	0	3350	NA	NA	0.26	NA	NA	0.26	NA	NA	0.01
												Sum of Deviations for All Items	0.53	0.41	0.71

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix F-3
Invariance of Item Parameters Across Years
Science Test

Item #	Used In Test Forms			Number of Responses			Proportion Correct for Item Respondents: Raw Item Responses			Proportion Correct for Item Respondents: IRT Estimates			Deviation (Actual-Predicted)		
				BY	F1	F2	BY	F1	F2	BY	F1	F2	BY	F1	F2
1	BY			23528	0	0	0.70	NA	NA	0.70	NA	NA	0.00	NA	NA
2	BY			23522	0	0	0.79	NA	NA	0.80	NA	NA	-0.01	NA	NA
3	BY	F1		23376	17583	0	0.65	0.73	NA	0.66	0.73	NA	-0.02	0.00	NA
4	BY	F1	F2	23464	17612	14070	0.67	0.75	0.79	0.69	0.75	0.78	-0.01	0.00	0.01
5	BY	F1	F2	23456	17611	14090	0.76	0.78	0.82	0.73	0.81	0.86	0.03	-0.03	-0.04
6	BY	F1	F2	23407	17607	14109	0.77	0.84	0.88	0.78	0.84	0.88	-0.01	0.00	0.00
7	BY			23403	0	0	0.66	NA	NA	0.63	NA	NA	0.03	NA	NA
8	BY			23514	0	0	0.57	NA	NA	0.57	NA	NA	0.01	NA	NA
9	BY			23498	0	0	0.65	NA	NA	0.65	NA	NA	0.00	NA	NA
10	BY	F1	F2	23225	17530	14034	0.54	0.60	0.66	0.51	0.61	0.67	0.03	-0.01	-0.01
11	BY			23086	0	0	0.49	NA	NA	0.49	NA	NA	0.00	NA	NA
12	BY	F1	F2	22341	17605	14088	0.70	0.73	0.73	0.65	0.73	0.77	0.05	0.01	-0.04
13	BY			22940	0	0	0.75	NA	NA	0.74	NA	NA	0.00	NA	NA
14	BY	F1	F2	23471	17600	14076	0.54	0.65	0.71	0.54	0.66	0.72	-0.01	0.00	-0.01
15	BY	F1	F2	23174	17506	13986	0.40	0.55	0.56	0.45	0.52	0.56	-0.05	0.03	0.01
16	BY	F1	F2	23157	17494	13853	0.47	0.56	0.59	0.47	0.55	0.60	0.00	0.02	-0.01
17	BY	F1	F2	23243	17520	13986	0.42	0.57	0.64	0.45	0.56	0.62	-0.02	0.02	0.02
18	BY	F1	F2	23160	17554	14017	0.46	0.58	0.65	0.47	0.57	0.63	-0.01	0.01	0.02
19	BY	F1	F2	23246	17584	14079	0.43	0.54	0.59	0.43	0.53	0.59	-0.01	0.01	0.00
20	BY	F1		23147	17516	0	0.42	0.50	NA	0.42	0.49	NA	0.00	0.01	NA
21	BY	F1		23149	17554	0	0.43	0.52	NA	0.43	0.51	NA	-0.01	0.01	NA
22	BY	F1	F2	22981	17378	13745	0.37	0.46	0.48	0.39	0.45	0.49	-0.02	0.01	-0.01
23	BY	F1		22613	17136	0	0.40	0.52	NA	0.43	0.48	NA	-0.02	0.04	NA
24	BY	F1	F2	23075	17477	13927	0.33	0.43	0.45	0.32	0.42	0.48	0.01	0.01	-0.02
25	BY	F1		22985	17442	0	0.22	0.33	NA	0.23	0.30	NA	-0.01	0.03	NA
26		F1	F2	0	17062	14030	NA	0.54	0.62	NA	0.54	0.60	NA	0.00	0.02
27		F1	F2	0	16754	13490	NA	0.29	0.33	NA	0.28	0.34	NA	0.01	-0.01
28			F2	0	0	14101	NA	NA	0.73	NA	NA	0.74	NA	NA	0.00
29		F1	F2	0	17347	14054	NA	0.49	0.58	NA	0.50	0.57	NA	-0.01	0.02
30		F1	F2	0	16745	13800	NA	0.52	0.60	NA	0.53	0.58	NA	-0.01	0.02
31			F2	0	0	14027	NA	NA	0.59	NA	NA	0.60	NA	NA	0.00
32		F1	F2	0	17089	13538	NA	0.26	0.36	NA	0.27	0.34	NA	-0.01	0.02
33		F1	F2	0	17263	13862	NA	0.57	0.65	NA	0.59	0.65	NA	-0.02	0.01
34		F1		0	17310	0	NA	0.47	NA	NA	0.46	NA	NA	0.01	NA
35			F2	0	0	13126	NA	NA	0.46	NA	NA	0.45	NA	NA	0.01
36			F2	0	0	13677	NA	NA	0.44	NA	NA	0.44	NA	NA	0.00
37			F2	0	0	13245	NA	NA	0.31	NA	NA	0.30	NA	NA	0.00
38			F2	0	0	13856	NA	NA	0.13	NA	NA	0.12	NA	NA	0.01
												Sum of Deviations for All Items	0.16	0.23	0.16

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix F-4
Invariance of Item Parameters Across Years
History/Citizenship/Geography Test

Item #	Used In Test Forms			Number of Responses			Proportion Correct for Item Respondents: Raw Item Responses			Proportion Correct for Item Respondents: IRT Estimates			Deviation (Actual-Predicted)		
				BY	F1	F2	BY	F1	F2	BY	F1	F2	BY	F1	F2
1	BY	F1	F2	23394	17527	14039	0.69	0.83	0.89	0.75	0.81	0.89	-0.06	0.02	0.00
2	BY	F1	F2	23237	17540	13982	0.49	0.65	0.66	0.51	0.59	0.71	-0.02	0.06	-0.05
3		F1		0	17539	0	NA	0.63	NA	NA	0.63	NA	NA	0.00	NA
4	BY	F1		23295	17512	0	0.48	0.56	NA	0.48	0.58	NA	0.00	-0.01	NA
5	BY	F1	F2	23377	17512	14003	0.55	0.68	0.71	0.58	0.64	0.74	-0.03	0.04	-0.03
6	BY	F1	F2	23088	17415	13858	0.43	0.51	0.55	0.40	0.48	0.61	0.03	0.03	-0.06
7	BY	F1		23485	17519	0	0.78	0.84	NA	0.79	0.84	NA	-0.01	-0.01	NA
8	BY	F1	F2	23265	17460	13998	0.59	0.68	0.77	0.58	0.67	0.78	0.00	0.01	-0.01
9	BY	F1	F2	23264	17484	13938	0.42	0.53	0.60	0.44	0.50	0.60	-0.01	0.03	0.00
10	BY	F1	F2	23443	17509	14003	0.47	0.52	0.62	0.44	0.52	0.66	0.03	0.00	-0.04
11	BY	F1	F2	23418	17501	14005	0.45	0.45	0.57	0.41	0.48	0.59	0.04	-0.03	-0.02
12			F2	0	0	13790	NA	NA	0.41	NA	NA	0.42	NA	NA	0.00
13	BY	F1	F2	23325	17468	13992	0.48	0.54	0.65	0.47	0.54	0.65	0.01	0.00	0.00
14	BY	F1		23433	17512	0	0.78	0.81	NA	0.78	0.83	NA	0.00	-0.02	NA
15	BY	F1	F2	23359	17515	14021	0.66	0.73	0.80	0.66	0.72	0.82	0.00	0.00	-0.02
16	BY	F1		23494	17508	0	0.90	0.91	NA	0.90	0.93	NA	0.00	-0.02	NA
17	BY	F1		23117	17376	0	0.81	0.86	NA	0.82	0.86	NA	-0.01	0.00	NA
18	BY	F1	F2	23063	17434	13797	0.25	0.29	0.58	0.27	0.34	0.48	-0.02	-0.05	0.10
19	BY	F1	F2	23406	17512	14042	0.84	0.91	0.96	0.87	0.91	0.95	-0.03	0.01	0.01
20			F2	0	0	13898	NA	NA	0.44	NA	NA	0.43	NA	NA	0.01
21	BY	F1	F2	23156	17445	13922	0.35	0.45	0.59	0.36	0.44	0.57	-0.01	0.01	0.02
22	BY			23292	0	0	0.87	NA	NA	0.87	NA	NA	0.00	NA	NA
23	BY			23311	0	0	0.85	NA	NA	0.84	NA	NA	0.01	NA	NA
24	BY			23302	0	0	0.92	NA	NA	0.92	NA	NA	0.00	NA	NA
25	BY			23299	0	0	0.89	NA	NA	0.89	NA	NA	0.01	NA	NA
26	BY			23308	0	0	0.92	NA	NA	0.92	NA	NA	0.00	NA	NA
27	BY	F1	F2	23395	17459	14040	0.77	0.81	0.91	0.76	0.83	0.91	0.01	-0.02	0.00
28			F2	0	0	13950	NA	NA	0.53	NA	NA	0.52	NA	NA	0.01
29	BY	F1		23325	17409	0	0.67	0.75	NA	0.68	0.75	NA	-0.01	0.00	NA
30	BY	F1		23425	17428	0	0.70	0.82	NA	0.73	0.79	NA	-0.03	0.03	NA
31	BY	F1	F2	23271	17451	14039	0.54	0.67	0.79	0.57	0.65	0.76	-0.03	0.02	0.02
32		F1	F2	0	17335	13814	NA	0.33	0.44	NA	0.33	0.43	NA	0.00	0.01
33	BY	F1	F2	22982	17273	13878	0.48	0.61	0.73	0.53	0.60	0.71	-0.05	0.01	0.03
34	BY	F1		23405	17359	0	0.59	0.55	NA	0.52	0.61	NA	0.08	-0.06	NA
35		F1		0	17362	0	NA	0.71	NA	NA	0.70	NA	NA	0.01	NA
36			F2	0	0	13449	NA	NA	0.26	NA	NA	0.26	NA	NA	0.01
37	BY	F1	F2	23151	17287	13927	0.52	0.57	0.69	0.49	0.58	0.71	0.03	-0.02	-0.02
38		F1		0	17221	0	NA	0.46	NA	NA	0.45	NA	NA	0.01	NA

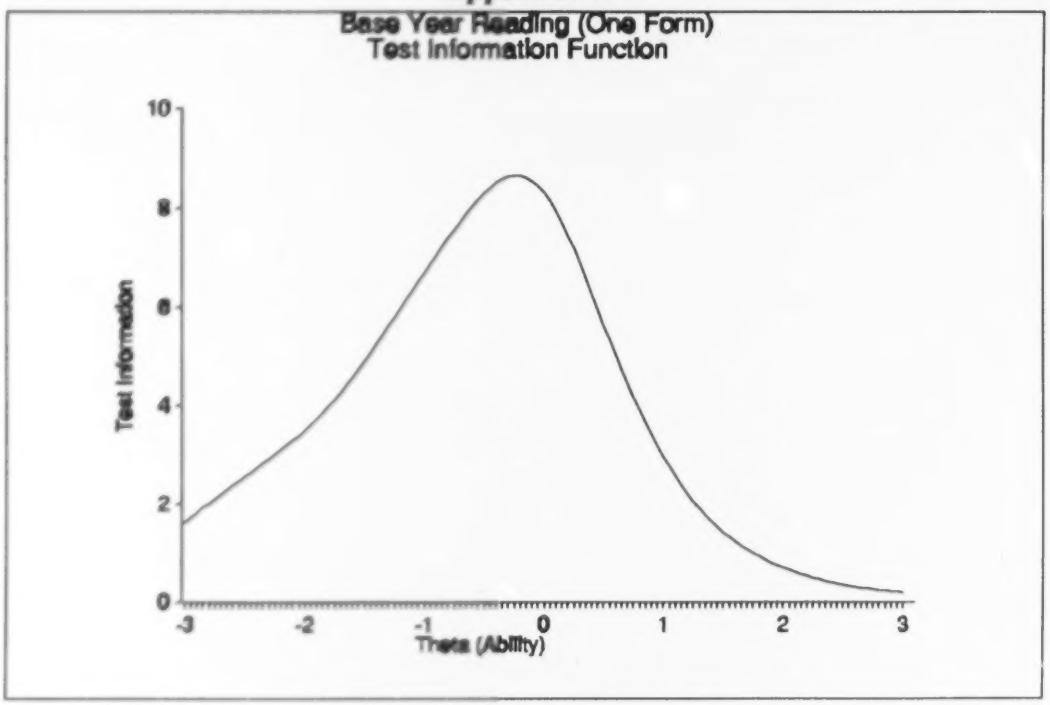
Appendix F-4
Invariance of Item Parameters Across Years
History/Citizenship/Geography Test (Continued)

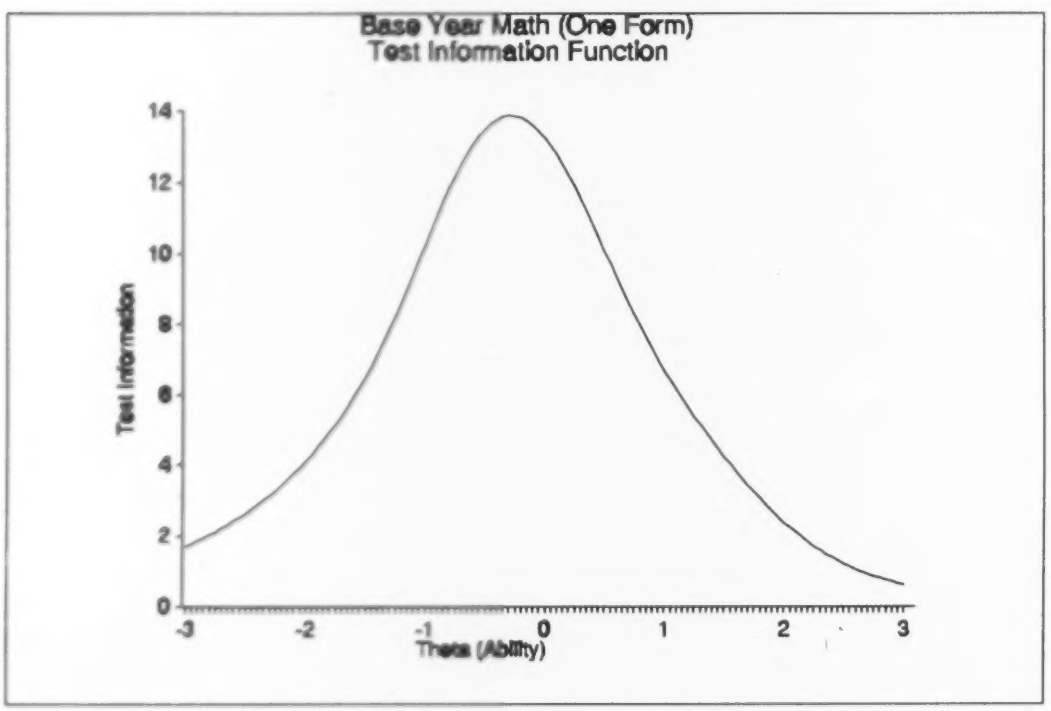
Item #	Used In Test Forms	Number of Responses			Proportion Correct for Item Respondents: Raw Item Responses			Proportion Correct for Item Respondents: IRT Estimates			Deviation (Actual-Predicted)		
		BY	F1	F2	BY	F1	F2	BY	F1	F2	BY	F1	F2
39	F1	0	17226	0	NA	0.42	NA	NA	0.41	NA	NA	0.01	NA
40	F2	0	0	13969	NA	NA	0.63	NA	NA	0.63	NA	NA	0.01
41	F2	0	0	13990	NA	NA	0.70	NA	NA	0.70	NA	NA	0.00
42	F2	0	0	13860	NA	NA	0.56	NA	NA	0.56	NA	NA	0.00
43	F2	0	0	13923	NA	NA	0.64	NA	NA	0.64	NA	NA	0.00
44	F2	0	0	13845	NA	NA	0.55	NA	NA	0.54	NA	NA	0.01
45	F2	0	0	13640	NA	NA	0.30	NA	NA	0.29	NA	NA	0.01
46	F2	0	0	13692	NA	NA	0.35	NA	NA	0.35	NA	NA	0.00
47	F2	0	0	13590	NA	NA	0.20	NA	NA	0.21	NA	NA	-0.01
								Sum of Deviations for All Items			0.24	0.29	0.23

Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix G: Test Information Function--Theta (Ability)

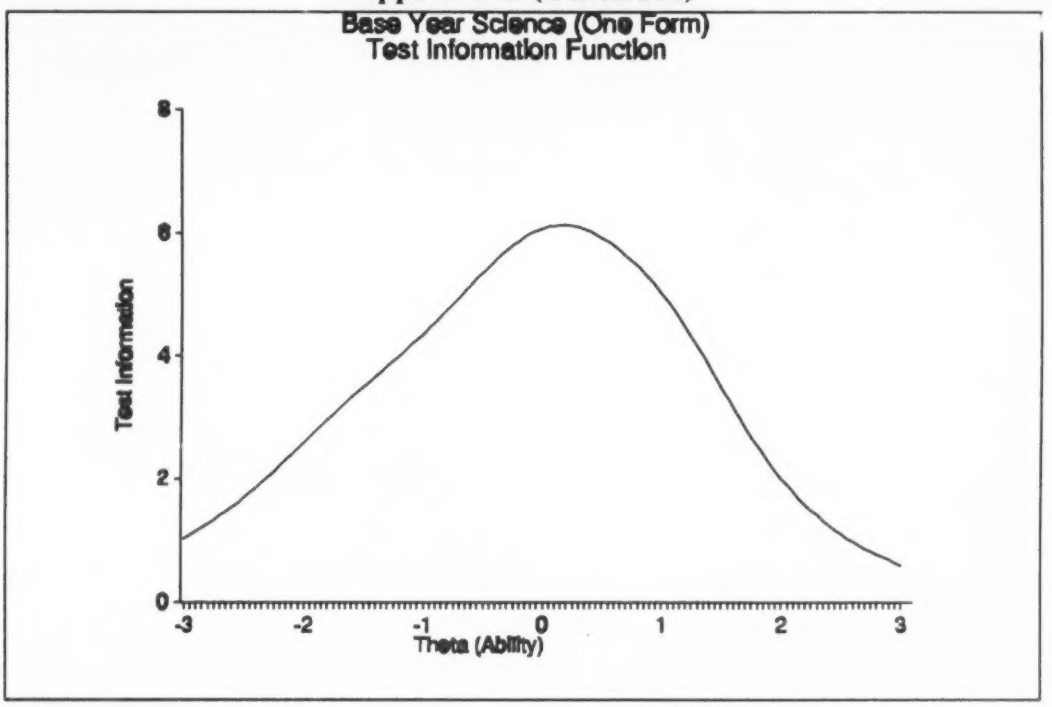
Appendix G

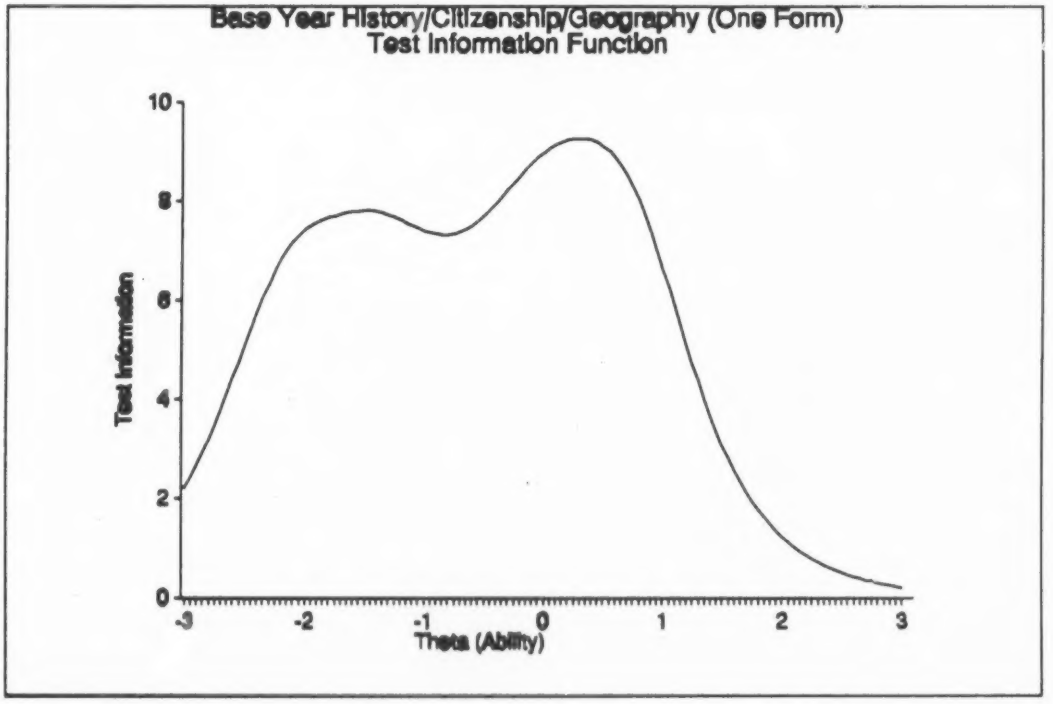




Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

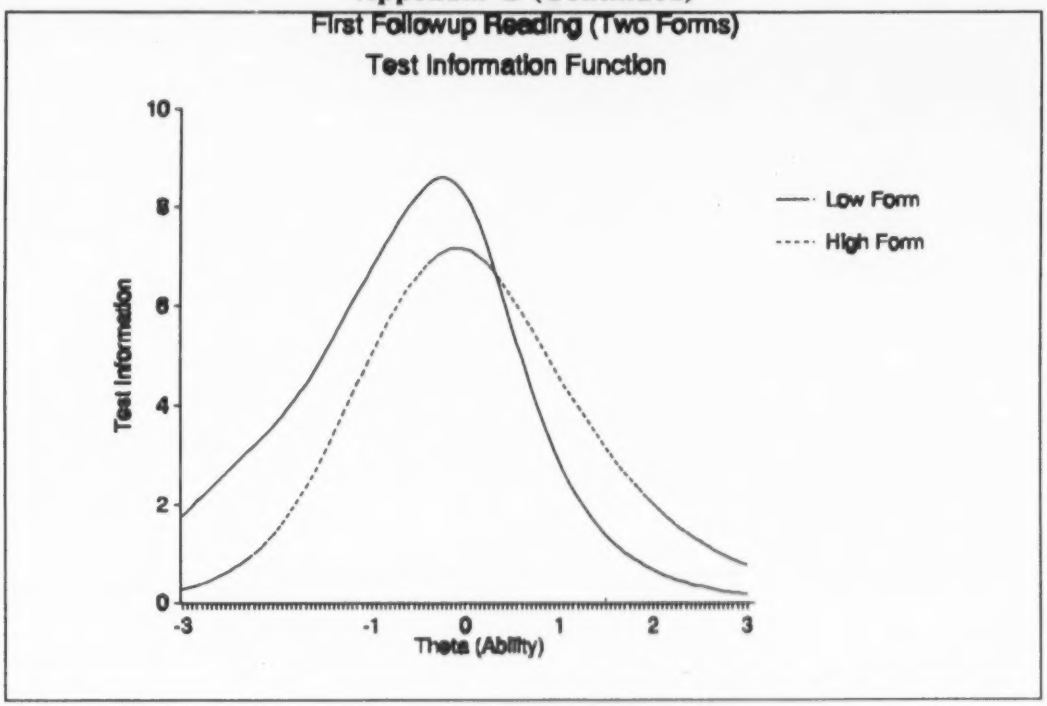
Appendix G (Continued)

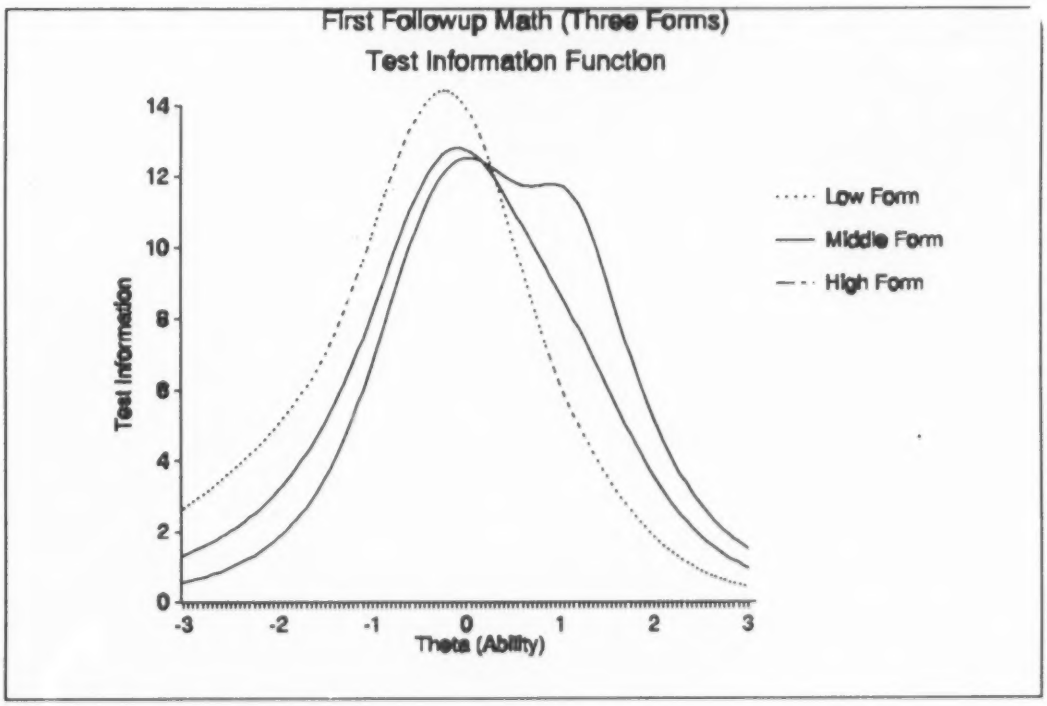




Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

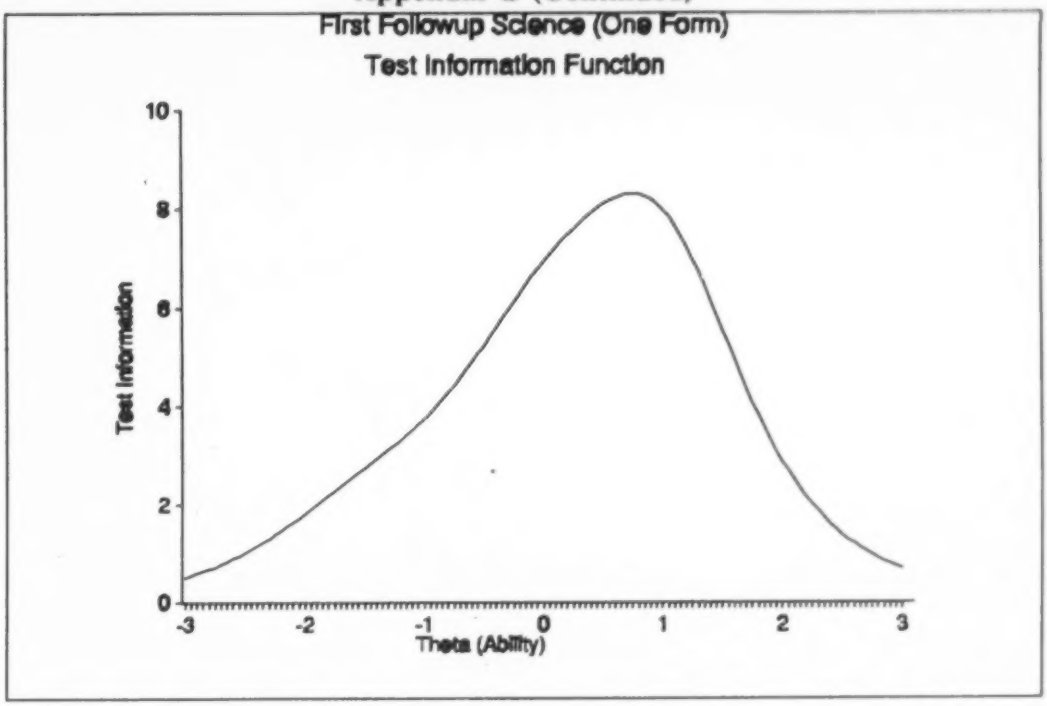
Appendix G (Continued)

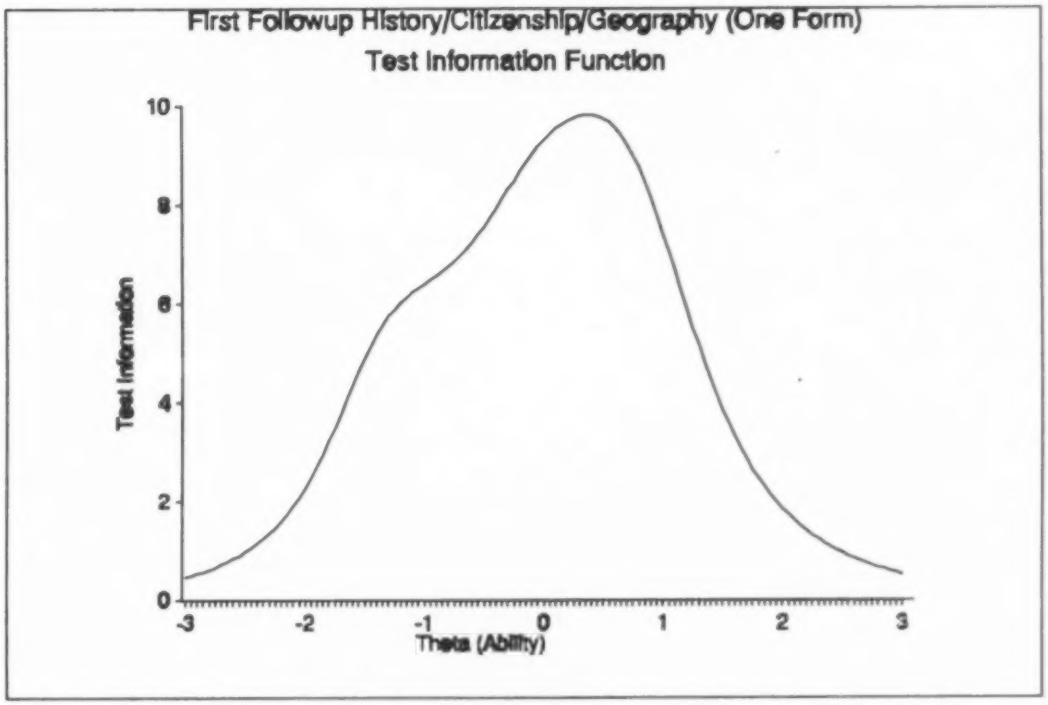




Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

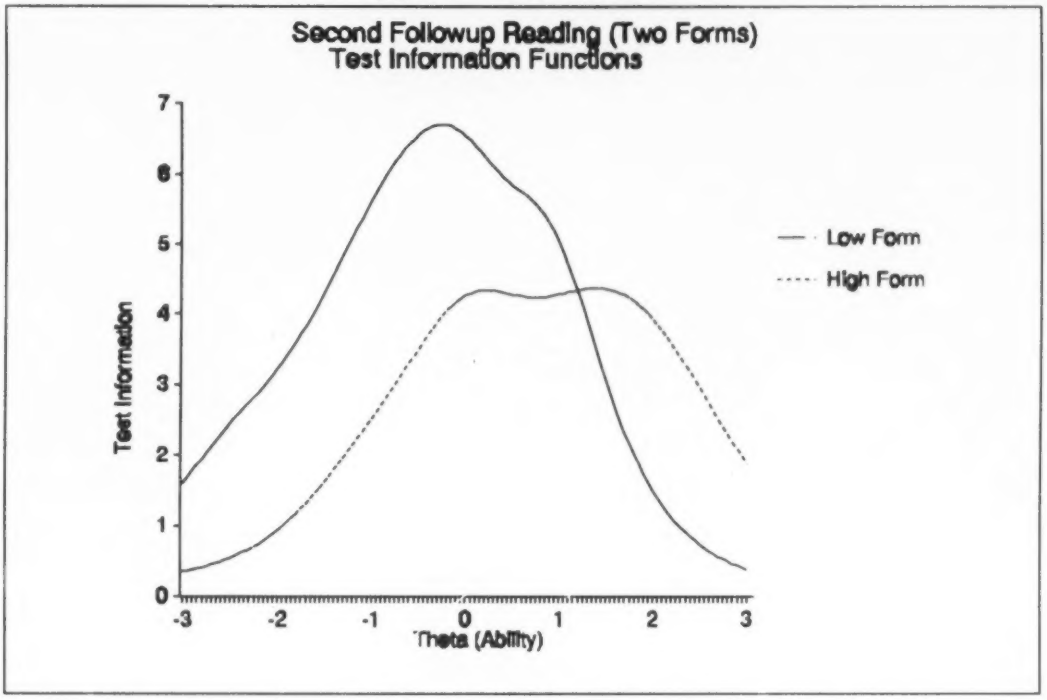
Appendix G (Continued)

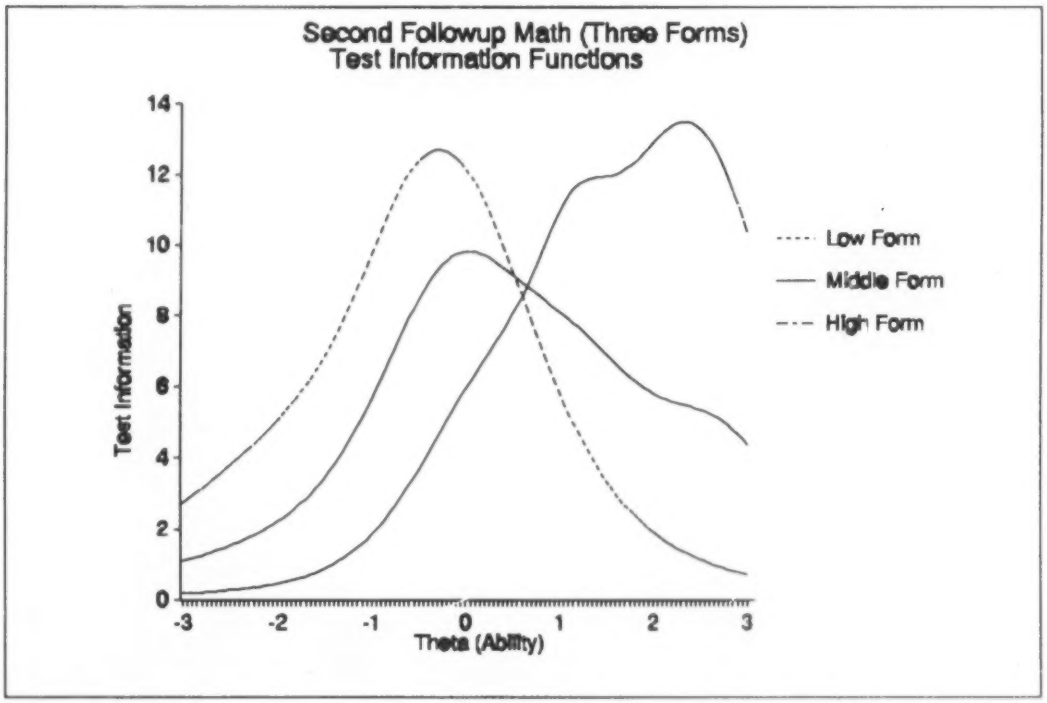




Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

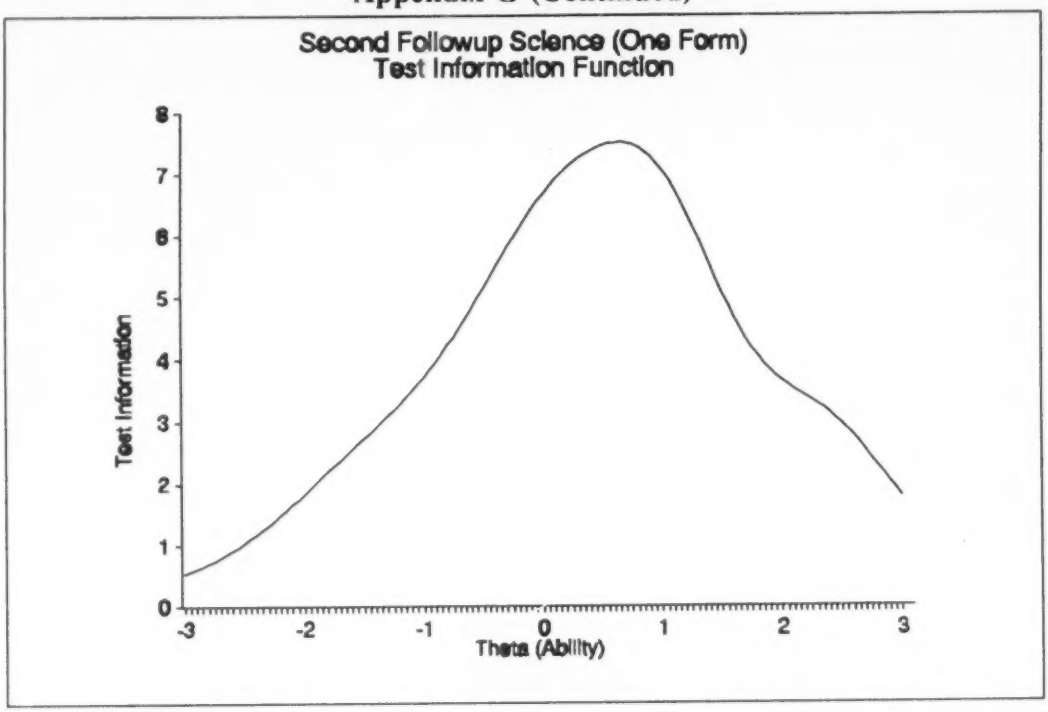
Appendix G (Continued)

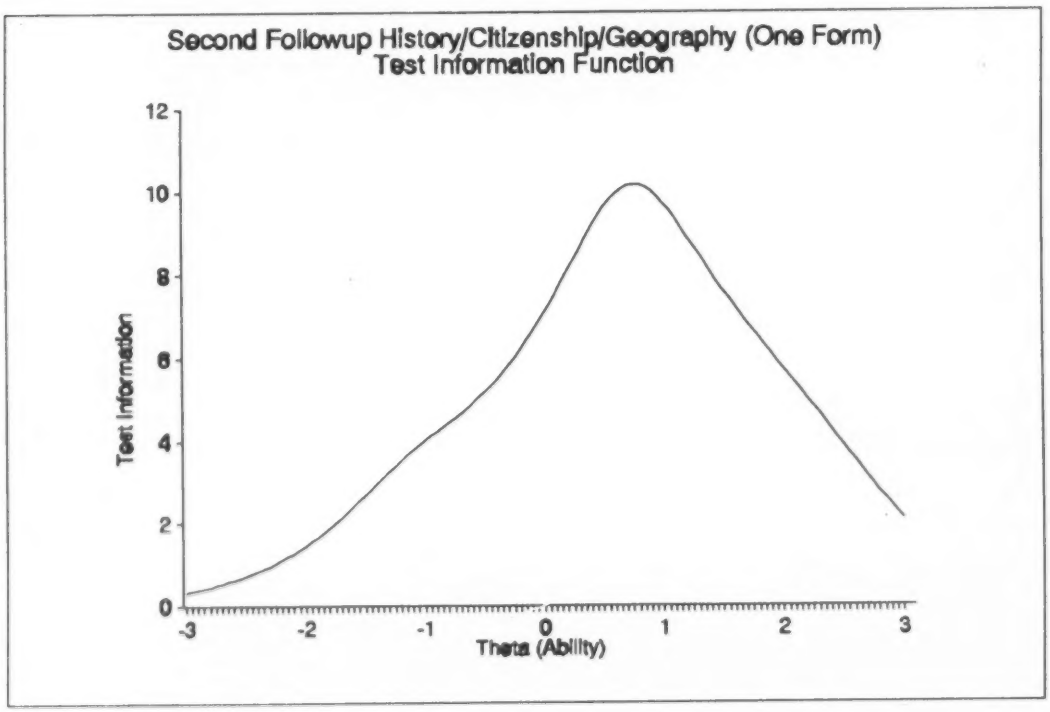




Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

Appendix G (Continued)





Source: National Education Longitudinal Study of 1988: Second Follow-Up, U.S. Department of Education, National Center for Education Statistics.

United States
Department of Education
Washington, DC 20208–5651

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Fourth Class Special
Special Handling